

Multiple Sclerosis Prevalence in New Zealand:
Effects of Latitude and UV

Submitted by Diane Patricia Fraser, to the University of Exeter
as a dissertation for the degree of
Master of Science by Research in Geography,
January 2012

This dissertation is available for Library use on the understanding that it is copyright material and that no quotation from the dissertation may be published without proper acknowledgement

I certify that all material in this dissertation which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University

Signature:

[This page intentionally left blank.]

Abstract

Multiple Sclerosis (MS) is a disease with a complex aetiology and pathogenesis. It is the most common neurologically debilitating disorder to affect young people, but it also affects older people. This dissertation is based on data from the New Zealand National MS Prevalence Study, a project funded by the New Zealand Health Research Council and the New Zealand MS Society which explores the relationship between the prevalence of MS in New Zealand, coincident with the national 2006 population census, and factors including gender, ethnicity, MS phenotype, latitude of residence and UV radiation exposure.

A latitudinal gradient of MS prevalence in New Zealand, which varies according to gender, ethnicity and MS phenotype, has been established previously. This dissertation extends this knowledge by examining the impact of lifetime residential migration in confounding this latitudinal gradient. In order to eliminate effects due to sample size, a prevalence–latitude rate ratio is considered for the prevalence of MS rather than the latitudinal gradient itself. Using GIS analysis, rate ratios have been determined for New Zealand. It is established that the north-south ratios differ according to gender, MS phenotype and case age at residence locations, with particular reference to early life locations. The female RR/SPMS ratio is higher than the male RR/SPMS ratio, however, the female PPMS ratio is lower than the male PPMS ratio. Early life location ratios appear similar to those at census prevalence date, however, differences between rate ratios for different factors are more significant.

The effect of replacing latitude with local ambient UV levels is also explored and the relationship between MS prevalence and UV exposure is seen to be very similar to that observed between MS prevalence and latitude.

The implications of these results for future studies to investigate possible causes of MS are discussed.

[This page intentionally left blank.]

Acknowledgements

I would like to thank my supervisor, Clive Sabel, for the opportunity to work on this project. In particular, I am indebted to him for his patience with me while I juggled the work and the writing with a full time job. I am also grateful to Bruce Taylor for many stimulating discussions, especially when we were able to meet in person. Thanks also go to the University of Exeter for sponsoring me through this study, and Geography and the College of Life and Environmental Sciences for allowing me to take on this project. Last, but by no means least, special thanks must go to my husband, Geoff, and to my parents, Ann and David, for their continued support and encouragement.

[This page intentionally left blank.]

Table of Contents

Abstract	3
Acknowledgements	5
Table of Contents	7
List of Figures.....	11
List of Tables	13
Abbreviations.....	15
Trademarks.....	16
Chapter 1 Introduction and Background Information	17
1.1 Background to MS and Regional Variations	17
1.2 Context of this Study	18
1.3 Aims and Objectives	18
1.4 Organisation of this Dissertation.....	19
Chapter 2 Multiple Sclerosis: A Complex Disease.....	21
2.1 Introduction.....	21
2.2 Aetiology, Pathology and Epidemiology.....	21
2.2.1 Clinical Features.....	24
2.2.2 Prevalence	26
2.2.3 Factors Affecting Prevalence	30
2.2.3.1 Gender and Familial.....	31
2.2.3.2 Ethnicity and Other Genetic Factors	33
2.2.3.3 Smoking	39
2.2.3.4 Epstein–Barr Virus	41
2.2.3.5 Diet	43
2.2.3.6 Latitude and Sunlight.....	46
2.2.4 Vitamin D—the Sunshine Vitamin.....	48
2.2.4.1 Photobiology, Photochemistry and Metabolism.....	48
2.2.4.2 Status and Maintaining Sufficiency	52
2.2.5 Timing of UV Exposure in MS	57

2.2.5.1	Prenatal.....	58
2.2.5.2	Childhood and Adolescence	59
2.2.5.3	Prior to Onset of MS	61
2.2.6	UV Exposure—Benefits and Risks.....	61
2.3	Summary of MS Epidemiology	63
Chapter 3 The Role of Geography in Understanding MS		65
3.1	Introduction.....	65
3.2	The Geography of MS	65
3.2.1	Genetics or Latitude?	67
3.2.2	MS Prevalence Studies Going Into the 21 st Century	71
3.3	The New Zealand National MS Prevalence Study	82
3.3.1	The NZMSPS Project	82
3.3.2	Prevalence of MS in New Zealand by Region	85
3.3.3	The Residence Calendar	87
3.3.4	Limitations of the NZMSPS Data.....	89
3.4	Areal Interpolation and the MAUP.....	90
3.4.1	The Modifiable Areal Unit Problem.....	90
3.4.2	Areal Interpolation	91
3.5	Migration	93
3.6	Summary of Geography.....	94
Chapter 4 Methodology—Data and Techniques.....		95
4.1	Introduction.....	95
4.2	Data Considerations for Different Life Periods	95
4.2.1	Residence Criteria for the Cohorts of MS Cases.....	95
4.2.2	Life Period Geographies	97
4.2.3	Census Boundaries and Population Data	100
4.2.3.1	The New Zealand 1956 Census Boundaries	101
4.2.3.2	The New Zealand 1956 Census Population Data.....	105
4.3	Analysis Techniques.....	108
4.3.1	GIS Techniques for Aggregating and Processing Data.....	109
4.3.1.1	Using ArcGIS for Areal Interpolation	110
4.3.1.2	ArcGIS ModelBuilder and Python Scripts	112

4.3.2	Age Standardisation of Data.....	112
4.3.2.1	Standard Populations	114
4.3.2.2	Calculating Age-Standardised Rates.....	115
4.3.3	Regression Techniques	116
4.3.3.1	Sample and Population Size Effects in Regression	118
4.3.4	Statistics and Curve Fitting Using R	119
4.3.5	Inter-Regional Migration	120
4.4	Summary of Methods	122
Chapter 5 Prevalence of MS with Latitude		123
5.1	Introduction.....	123
5.2	Overview of Prevalence of MS in New Zealand	124
5.2.1	Consequences of Areal Interpolation.....	129
5.2.2	Choice of Standard for Age Standardisation	143
5.3	Effect of Factors on Prevalence	146
5.3.1	Gender and MS Phenotype	146
5.3.2	Migration—External	163
5.3.3	Migration—Internal.....	167
5.3.4	Month of Birth	173
5.4	Summary of Latitude Results.....	178
Chapter 6 Variations in the Prevalence of MS with UV Exposure.....		181
6.1	Introduction.....	181
6.2	Variations in Solar and UV Radiation.....	181
6.2.1	Insolation and UV Irradiance	182
6.2.2	Action Spectra	184
6.2.3	NIWA UV Data	188
6.2.3.1	Surface Plots	191
6.2.3.2	Time Series	192
6.3	Average UV Exposure with Latitude.....	198
6.4	MS Prevalence and UV—Discussion of Analysis.....	201
6.5	Summary of UV Exposure	204
Chapter 7 Summary and Conclusions		205
7.1	Introduction.....	205

7.2	Summary and Main Findings	205
7.2.1	General Prevalence.....	206
7.2.2	Prevalence–Latitude Gradient and Migration.....	207
7.2.3	Month of Birth	210
7.3	Implications of Results.....	211
7.3.1	UV Exposure and Vitamin D.....	211
7.3.2	Prevalence of MS, Latitude Gradients, and Migration	214
7.4	Strengths and Limitations of the Study	216
7.5	Further Research	217
7.6	Conclusions	218
Appendix A	The NZMSPS Questionnaire.....	219
Appendix B	1956 Census Data	231
B.1	Constructing the 1956 Census Boundaries	231
B.2	Tables for Determining Census Populations.	234
Appendix C	ArcGIS Python Scripts	245
C.1	Processing with ArcGIS ModelBuilder and Python.....	245
C.2	ArcGIS Python Scripts	250
Appendix D	Least Squares Regression	321
D.1	Linear Least Squares Regression	321
D.2	Log and Logistic Least Squares Regression.....	324
Appendix E	Migration Models	329
Appendix F	Solar Insolation	333
References	337

List of Figures

Figure 2.1	A schematic of a neuron and demyelination	23
Figure 2.2	The pathogenesis of MS	24
Figure 2.3	Progression of the severity of symptoms of MS	25
Figure 2.4	Factors affecting the prevalence of MS.....	27
Figure 2.5	The European family tree	34
Figure 2.6	T-cell biological pathways associated with a susceptibility to MS.....	38
Figure 2.7	Photosynthesis of vitamin D in the skin	49
Figure 2.8	Penetration of incident sunlight to different depths of the skin	50
Figure 2.9	Schematic of neural crest formation.....	58
Figure 3.1	Age-adjusted death rates for MS by country	66
Figure 3.2	Age-adjusted prevalence rates for MS by WHO member state.....	79
Figure 3.3	MS in Australia and New Zealand.....	81
Figure 3.4	Aggregated regions in New Zealand.....	86
Figure 3.5	Prevalence of MS in New Zealand 2006.....	87
Figure 3.6	The Residence Calendar which MS cases were asked to fill out.....	88
Figure 4.1	Distributions of years of birth and onset, and age at onset of MS	99
Figure 4.2	Geographic Hierarchy for the 2006 NZ Census	101
Figure 4.3	1956 census areas of North Island New Zealand	103
Figure 4.4	1956 census areas of South Island New Zealand	104
Figure 4.5	Areal interpolation of New Zealand 2006 census areas.....	111
Figure 4.6	The process and data flow for the analyses.....	113
Figure 4.7	European and Segi World standard populations	115
Figure 4.8	Modelling inter-regional migration	121
Figure 5.1	Variation in prevalence of MS in New Zealand—census day 2006.....	127
Figure 5.2	Variation in prevalence of MS in New Zealand—Age 0–4	128
Figure 5.3	MS case numbers and population—mean latitudes.....	130
Figure 5.4	MS case numbers and population—population weighted latitudes ...	131
Figure 5.5	Linear regression fits—mean latitudes for 2006 Prevalence	134
Figure 5.6	Log regression fits—mean latitudes for 2006 Prevalence	135
Figure 5.7	Linear regression fits—weighted latitudes for 2006 Prevalence	136
Figure 5.8	Log regression fits—weighted latitudes for 2006 Prevalence	137
Figure 5.9	Linear regression fits—weighted latitudes for Age 0–4 Prevalence	138

Figure 5.10	Log regression fits—weighted latitudes for Age 0–4 Prevalence	139
Figure 5.11	Rate Ratios—Latitude Slice Size	142
Figure 5.12	Variation of age structure with Latitude	144
Figure 5.13	Log regression fits—2006 Prevalence by Gender	148
Figure 5.14	Log regression fits—Age 0–4 Prevalence by Gender	149
Figure 5.15	Rate Ratios—Gender and Life Period	151
Figure 5.16	Log regression fits—2006 Prevalence by MS Phenotype (Female).....	153
Figure 5.17	Log regression fits—Age 0–4 Prevalence by MS Phenotype (Female).	154
Figure 5.18	Log regression fits—2006 Prevalence by MS Phenotype (Male)	155
Figure 5.19	Log regression fits—Age 0–4 Prevalence by MS Phenotype (Male)	156
Figure 5.20	Rate Ratios—Life Period, Gender and MS Phenotype	159
Figure 5.21	Rate Ratios—Gender and MS Phenotype (2006)	160
Figure 5.22	Rate Ratios—Gender and MS Phenotype (Age 0–4)	161
Figure 5.23	Log regression fits—2006 Prevalence for Migrated cohort	165
Figure 5.24	Net migration within New Zealand—Pre Onset.....	171
Figure 5.25	Net migration within New Zealand—Post Onset	172
Figure 5.26	Number of births in New Zealand by month of birth.....	174
Figure 5.27	Crude prevalence by Month of Birth.....	175
Figure 6.1	Simulated UVB radiation—solar elevation and cloud	184
Figure 6.2	Action Spectra and Solar Radiation	186
Figure 6.3	NIWA Weather Stations	189
Figure 6.4	Spatial Variations in Clear-Sky UV by Month	193
Figure 6.5	NIWA Weather Stations (22)—Locations	194
Figure 6.6	Time Series—Monthly Totals	197
Figure 6.7	UV Exposure – Latitude	199
Figure 6.8	Prevalence of MS—UV Exposure.....	200
Figure B.1	Constructing the 1956 census areas from 2006 meshblocks.....	232
Figure B.2	Constructing the 1956 census areas for Auckland	233
Figure C.1	A simple ArcGIS ModelBuilder model for processing census data	245
Figure C.2	An ArcGIS ModelBuilder model with parameters	246
Figure F.1	Annual Variations in Relative Solar Radiation by Latitude	335
Figure F.2	Daily Variations in Relative Solar Radiation by Latitude	336

List of Tables

Table 2.1	Selected prevalence rates in relatives of MS probands	33
Table 2.2	Blood serum levels of 25(OH)D for vitamin D status	53
Table 3.1	MS prevalence publications—studies in Italy	72
Table 3.2	MS prevalence publications—studies in Europe, excluding Italy	73
Table 3.3	MS prevalence publications—studies in non-European countries	75
Table 3.4	MS prevalence publications—studies covering several countries.....	77
Table 3.5	NZMSPS project investigators	83
Table 3.6	A summary of Section 1 of the NZMSPS questionnaire	84
Table 3.7	Summary findings of MS phenotype from the NZMSPS	85
Table 4.1	1956 Census Tables	105
Table 4.2	Data available from the 1956 Census: Hardcopy and Electronic	106
Table 4.3	Bay of Islands population data from the 1956 Census.....	107
Table 5.1	List of Tables in Chapter 5	123
Table 5.2	List of Figures in Chapter 5	124
Table 5.3	Prevalence of MS with latitude in New Zealand	125
Table 5.4	Gradients of prevalence of MS with latitude in New Zealand	128
Table 5.5	Linear and log regression gradients—mean latitudes	140
Table 5.6	Linear and log regression gradients—weighted latitudes.....	141
Table 5.7	Linear and log regression gradients—Segi Standard.....	145
Table 5.8	Numbers of cases of MS by Gender and MS Phenotype	146
Table 5.9	Linear and log regression gradients—Gender.....	150
Table 5.10	Linear and log regression gradients—MS Phenotype (Female).....	157
Table 5.11	Linear and log regression gradients—MS Phenotype (Male)	158
Table 5.12	Rate Ratio Comparisons for PPMS Phenotype and 3° latitude slices... 163	
Table 5.13	Log regression gradients—2006 Prevalence for Migrated cohort	166
Table 5.14	Net Migration—Gender.....	168
Table 5.15	Net Migration—Gender and MS Phenotype	169
Table 6.1	NIWA UV Atlas Program—Data Inputs.....	190
Table 6.2	NIWA UV Atlas Program—Weighting Functions	190
Table 6.3	The NIWA UV Atlas Program—Data Outputs.....	191
Table 6.4	NIWA Weather Stations (22)—Data Availability	195
Table 7.1	Percentage rates of change of prevalence with latitude	208

Table C.1	The naming convention used for the output data	248
Table C.2	List of python scripts	250
Table D.1	ANOVA table	323

Abbreviations

ALS	amyotrophic lateral sclerosis	MZ	monozygotic
AU	areal unit	NIWA	National Institute of Water and Atmospheric Research
BED	biologically effective dose	NZ	New Zealand
CAU	census areal unit	NZMSPS	NZ National MS Prevalence Study
CI	confidence interval	OSMS	optic spinal MS
CMS	conventional MS	PNS	peripheral nervous system
CNS	central nervous system	pop	population
CIDP	chronic inflammatory demyelinating polyradiculoneuropathy	PPMS	Primary Progressive MS
CNS	central nervous system	PRMS	Progressive Relapsing MS
DBP	vitamin D binding protein	PWC	population weighted centroid
DMS	definite multiple sclerosis	RC	Regional Council
DNA	deoxyribonucleic acid	RRMS	Relapsing-Remitting MS
DZ	dizygotic	RR/SPMS	RRMS/SPMS
EBV	Epstein–Barr virus	SDD	standard vitamin D dose
F:M	female to male	SED	standard erythemal dose
GIS	geographical information system	SN	Azzalini skew normal
GWAS	genome-wide association study	SNP	single nucleotide polymorphism
HERV	human endogenous retrovirus	SPF	sun protection factor
HLA	human leukocyte antigen	SPMS	Secondary Progressive MS
IFN	interferon	StatsNZ	Statistics New Zealand
IL	interleukin	TA	Territorial Authority
IM	infectious mononucleosis	TLA	Territorial Local Authority
IR	infrared	TNF	tumor necrosis factor
IU	international unit	UCA	urocanic acid
MAUP	modifiable areal unit problem	UK	United Kingdom
MB	meshblock	US	United States
MED	minimal erythemal dose	UV	ultraviolet
MHC	major histocompatibility complex	UVA	ultraviolet-A
MS	multiple sclerosis	UVB	ultraviolet-B
MSRV	MS-associated retrovirus	UVC	ultraviolet-C
		UVR	ultraviolet radiation
		VDR	vitamin D receptor
		WHO	World Health Organization

Trademarks

ArcCatalog, ArcGIS, ArcMap, ArcObjects, ArcToolbox, Esri and ModelBuilder are trademarks, registered trademarks, or service marks of Esri in the United States, the European Community, or certain other jurisdictions.

Linux is a registered trademark of Linus Torvalds in the United States and other countries.

Microsoft, Access, Excel, and Windows are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

S and S-Plus are trademarks or registered trademarks of Insightful Corporation in the United States and/or other countries.

Chapter 1

Introduction and Background Information

1.1 Background to MS and Regional Variations

Multiple Sclerosis (MS) is a complex neurological disease, with unknown aetiology. The prevalence rates of MS vary considerably according to regional location, ethnicity and gender, amongst other factors. Understanding how these factors interact, and how individual factors affect the prevalence rates in the absence of other confounding factors, is key to understanding MS. In particular, females of white European ancestry are seen to be at greatest risk, and variations in prevalence rates with location tend to be associated with latitude of residence, with latitude thought to be a surrogate for UV exposure, possibly linked to vitamin D synthesis in the body. The so called 'latitude effect' has been observed in many countries in both hemispheres, including in Australia (Taylor et al. 2010a), France (Vukusic et al. 2007) and Sweden (Ahlgren et al. 2011). The link with UV exposure and vitamin D is not only of concern in the study of MS, it is also a current topic of major concern in the increasing prevalence of rickets (Holick 2006) and other diseases such as diabetes and many internal cancers, in addition to osteomalacia, cardiovascular disease, rheumatoid arthritis and influenza (Juzeniene et al. 2011). There are also commonalities between MS and many auto-immune diseases with shared genetic and biological pathways (Baranzini 2011).

New Zealand is a suitable location for studying the prevalence of MS for a number of reasons. The North, South and Stewart Islands of New Zealand cover a range of 13° of latitude from approximately 34.4°S to 47.3°S, and age-standardised prevalence rates of MS for New Zealand have previously been reported as ranging from 29.6 per 100,000 in the north of the country to 81.7 per 100,000 in the south (Miller et al. 1990). This also makes New Zealand ideal for studying variations in the prevalence of MS with latitude. The population of New Zealand was 4.1 million in 2006; this is small enough to allow for a country-wide survey, yet large enough to provide adequate sample sizes for subsequent analysis. 51% of the New Zealand population is female, and most of the New Zealand population is of white European ancestry, with just under 14% of Māori ethnicity, so the effects of gender and ethnicity on the prevalence of MS can be examined. (Note that here, and throughout this dissertation, gender is used to refer to biological sex at birth.)

1.2 Context of this Study

This study forms a part of a larger project, and, as such, there has been a limit to some of the analyses which could be carried out; where this is the case, this will be indicated in this dissertation. The larger project, the New Zealand National MS Prevalence study (NZMSPS), focused around a survey taken in conjunction with the New Zealand population census, 7 March 2006. The data analysed within this study was taken from the NZMSPS. The key findings from the NZMSPS have been published by Taylor et al (2010b).

1.3 Aims and Objectives

The principal aim of this study was to investigate the spatial variation of the prevalence of MS in New Zealand (NZ) with particular reference to gender, MS phenotype and past residence locations to establish how the prevalence of MS varies with the latitude of residence of MS cases within New Zealand at various periods of life. Although some initial prevalence results had already been obtained from the NZMSPS data, this study expands on the analyses and extends the results.

The specific objectives of this study were to:

- Determine the effects on the prevalence rates of MS obtained using different analysis techniques
- Compare the effects of selecting subgroups of cases with results from the whole sample of cases
- Compare the prevalence rates obtained for different periods of life
- Determine the effects of stratification by gender and MS phenotype
- Investigate the link between UV exposure and latitude with reference to prevalence of MS

As part of this study, techniques were developed to incorporate historical NZ census population data into the analyses. In addition, UV exposure data was examined and techniques explored which will assist future investigations into the variations of MS prevalence with UV exposure of MS cases.

1.4 Organisation of this Dissertation

This dissertation comprises seven chapters:

- Chapter 1 provides a brief introduction to MS, the importance of understanding variations in the prevalence of MS and the suitability of New Zealand as a study location. It continues with the aims of this study;
- Chapter 2 extends the introduction to MS with an outline of the aetiology, pathology and epidemiology of MS. The role of the immune system in MS and current understanding of vitamin D and the various roles of UV radiation are also detailed;
- Chapter 3 looks at the geography of MS and at the New Zealand National MS Prevalence Study which underpins this study, and which provides the data being analysed here. Prevalence results based on regional aggregates are summarised and limitations of the data are discussed. This chapter also describes some of the problems with data which varies with time and geography and introduces migration;
- Chapter 4 extends the analysis of prevalence from Chapter 3 to include past residence data. The use of GIS in enabling the aggregation of the data by latitude rather than by regions is described along with various analysis techniques;
- Chapter 5 presents the main results from this study. The variations in the prevalence of MS with latitude are examined and effects of the analysis techniques used in this study are investigated;
- Chapter 6 examines the efficacy of UV and considers spatial and temporal variations in UV radiation exposure. UV data from NIWA is explored and the incorporation of UV data into future MS prevalence studies is discussed;
- Chapter 7 summarises this study with a discussion of the main findings and suggestions for further work. Implications of these findings are also discussed.

[This page intentionally left blank.]

Chapter 2

Multiple Sclerosis: A Complex Disease

2.1 Introduction

Although there are many excellent reviews in the literature (Ascherio and Munger 2007a, 2007b; Barnett et al. 2009; Compston and Coles 2008; Ebers and Goodin 2007; Koch-Henriksen and Sørensen 2010; Lassmann 1999; Maghzi et al. 2011a; Murray 2009; Pugliatti et al. 2006; Rosati 2001; Sellner et al. 2011; Solomon and Whitham 2010; Sospedra and Martin 2005), this chapter gives an overview of Multiple Sclerosis (MS) in order to provide background information pertinent to the rest of this dissertation. Elements of the pathology and clinical features are described along with aspects of the immune system which are key to understanding MS. The role of ultra violet (UV) radiation exposure and the balance between its dangers and benefits are examined with a summary of current understanding.

2.2 Aetiology, Pathology and Epidemiology

The first well recorded case of MS is probably that of Sir Augustus Frederick d'Esté, a grandson of George III. In 1941 Firth reported the finding of a diary written by d'Esté between 1822 and 1846:

"He gives a vivid account of his own complaint, disseminated sclerosis, the clinical picture including transitory blindness and diplopia, absolute loss of power of locomotion followed by almost complete recovery, so that " the Chamois hunting " could be followed later, the gradual onset of paraplegia, tremors, sphincter troubles and sensory symptoms."
(Firth 1941, p.381)

However, three hundred years earlier, in preparation for her possible canonisation, the Vatican documented the Dutch case of Saint Ludwina of Schiedam (Murray 2009), the patron saint of ice skating. From the early age of sixteen she suffered from a variety of symptoms which we would now associate with multiple sclerosis, though at the time, her illness, and recovery between relapses, was attributed as coming from God. In 1868 Jean Marie Charcot described *La Sclérose en Plaques* (Landtblom et al. 2010; Lassmann 1999; Murray 2009), the first description of a disease identifiable as MS. MS was also illustrated in several atlases of pathology, notably those of Sir Robert Carswell prior to 1837 and Jean Cruveilhier between 1829 and 1842 (Landtblom et al.

2010; Lassmann 1999; Murray 2009). Yet despite being known of for almost two centuries, the aetiology of MS is still unknown.

Until recently MS was widely believed to be an autoimmune disease (Lucas et al. 2011b; Pender and Greer 2007; Ponsonby et al. 2005; Sospedra and Martin 2005; Youinou et al. 2009) and it is well established that the immune system is involved in its pathology (Arnason 2011; Hemmer et al. 2006; Krone et al. 2011; Rauchway 2011; Smolders et al. 2008; Solomon and Whitham 2010), however, results of research into MS cannot satisfy the criteria for an autoimmune disease (Behan and Chaudhuri 2010; Behan 2010; Ebers 2009; Rodriguez 2009) and researchers are now beginning to reassess their whole approach to understanding the disease, from prevention, through treatment and, hopefully, even to providing a cure (Bains 2010; Fatima et al. 2011; Giovannoni 2011; Goodin 2009; Hanwell and Banwell 2011; Kakalacheva and Lünemann 2011; Krone and Grange 2011; Lünemann and Münz 2009; Maghzi et al. 2011b; Mehta 2010; Pender 2011). Despite the elusive nature of the MS aetiology, there are many things which are known about MS. A detailed examination of the biochemistry involved is beyond the scope of this study, as well illustrated by several authors including Figure 4 of Sospedra and Martin (2005) which gives a comprehensive "Schematic diagram depicting the pathogenetic steps and contributing factors that lead to tissue damage in MS". However, an outline is now given to enable a better understanding of the epidemiology and pathology of MS as a background to this study.

MS takes its name from the scleroses, plaques or lesions, which form in the brain and on nerve cells within the central nervous system (CNS). However, the main damage which is observed is to the myelin sheaths which insulate nerves from the surrounding tissue as shown in Figure 2.1. Key to myelinated nerve function is the action potential which is set up between adjacent nodes of Ranvier. Demyelination in itself does not totally disrupt nerve signals, it just changes their speed of conductance (Rushton 1951), and many vertebrates have nerves which are not myelinated. Instead, it is likely that the demyelination causes chemical imbalances within the surrounding tissues, and it is these chemical imbalances which disrupt normal nerve function enough to cause the symptoms experienced by sufferers of MS and other demyelination diseases (Coggan et al. 2010). The variety of different symptoms experienced by sufferers is also consistent with the variety of chemical imbalances which could be experienced.

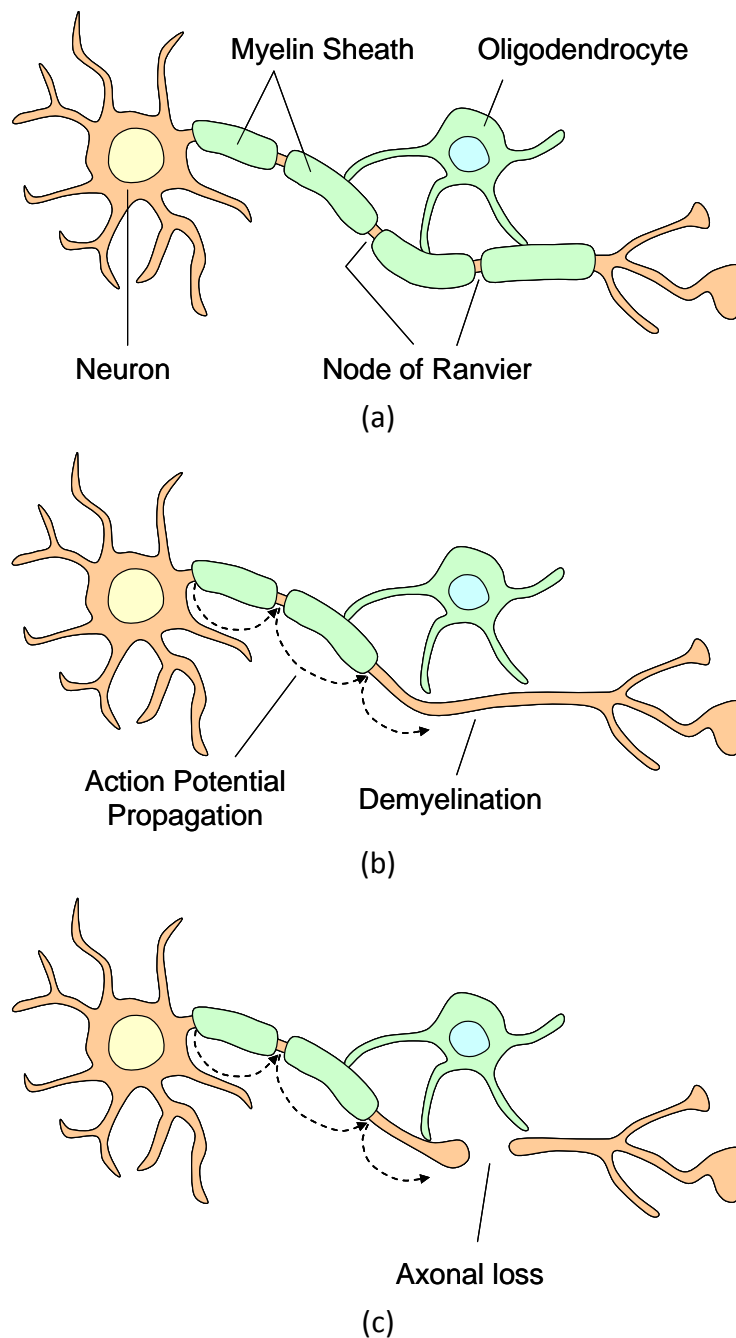


Figure 2.1 A schematic of a neuron and demyelination

(a) An intact neuron showing the myelin sheaths with the nodes of Ranvier, and an oligodendrocyte cell, the CNS equivalent of the schwann cells found in the peripheral nervous system (PNS). **(b)** A neuron exhibiting demyelination. Propagation of nerve signals via the action potentials between the nodes of Ranvier is interrupted. The oligodendrocyte cell may be able to repair the demyelination and restore nerve function. **(c)** A neuron exhibiting axonal loss. Nerve function damage is permanent.

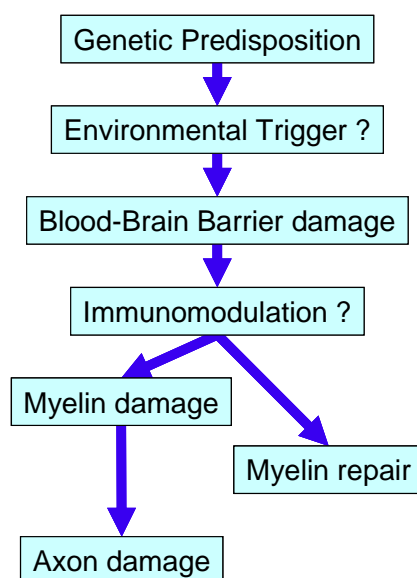


Figure 2.2 The pathogenesis of MS

Exactly what the environmental trigger is, or exactly what causes the immunomodulation is still unknown. However, each is likely to be a combination of factors rather than any single factor.

Damage caused by nerve cell demyelination can be repaired (Kornek and Lassmann 2003; Moore and Esiri 2011), though, as yet, not permanently (Patel and Klein 2011; Stangel et al. 2011), and it is the changing balance between damage and repair which appears to lead to the different groups of symptoms experienced as outlined in the next section. Exactly what causes the original lesions and triggers the subsequent nerve damage is the main topic of research in MS and is still not fully understood (Dutta and Trapp 2011; Lassmann 2002; McQualter and Bernard 2007; Schirmer et al. 2011; Sriram 2011), though the overall pathogenesis is believed to be as illustrated in Figure 2.2. Some of the factors involved are discussed later in this chapter.

2.2.1 Clinical Features

Individually, none of the symptoms experienced by sufferers of MS are unique to MS, and the patient can experience any or all of the symptoms without being diagnosed as having MS. Diagnosis of MS, therefore, is not trivial, though one of the main criteria is the repeated nature of the attacks (Mallam and Scolding 2009). Symptoms can include, in various combinations: double vision and other problems with sight including, colour vision and visual acuity; clumsiness or tremour; vertigo; bowel, bladder and erectile dysfunction; motor symptoms such as fatigue, stiffness or weakness; sensory symptoms such as tingling, burning or loss of sensation; headaches and migraines;

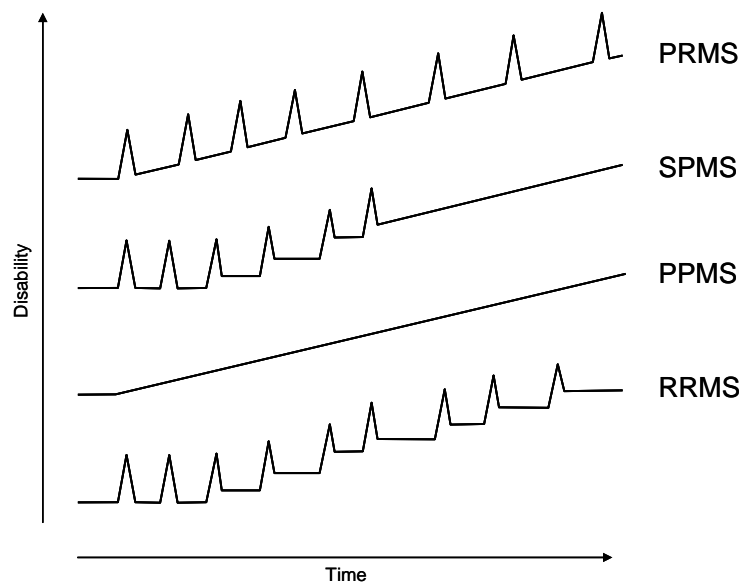


Figure 2.3 Progression of the severity of symptoms of MS

The four phenotypes are illustrated with spikes indicating attacks or relapses followed by remission. Note that, often, remission does not always return the disability level back to the same value as before the attack.

language skills and memory loss or lack of concentration; mood disorders including depression (Compston and Coles 2008; Maghzi et al. 2011a).

Once someone has exhibited symptoms of MS, the severity and frequency of the symptoms tend to follow one of four trends. These are illustrated in Figure 2.3 and define the four phenotypes of MS:

RRMS: Relapsing-Remitting MS:

- 80–85% of patients initially present with RRMS.
- Unpredictable attacks or relapses with months or years of remission.
- Damage caused during relapses may or may not be fully repaired.

SPMS: Secondary Progressive MS:

- 50–65% of RRMS sufferers progress to SPMS within 10–15 years.
- Progressive neurological decline with little or no remission.

PPMS: Primary Progressive MS:

- 10–15% of patients initially present with PPMS.
- Progressive neurological decline from initial attack with little or no remission.

PRMS: Progressive Relapsing MS:

- Progressive neurological decline from initial attack with additional distinct attacks.

Although RRMS and SPMS are separate phenotypes, a majority of RRMS sufferers progress to SPMS (Rovaris et al. 2006), so the two phenotypes are often grouped together (RR/SPMS). This avoids problems with age standardisation effects since the age profile of RRMS sufferers will otherwise be skewed to lower ages and the age profile of SPMS sufferers will be significantly skewed to higher ages. (Age standardisation will be discussed in Chapter 4.) Combining the two phenotypes also acknowledges the fact that SPMS can be considered as a later stage of RRMS.

Note that although the above phenotypes are those traditionally associated with MS, studies in Asia, and in particular Japan, have suggested that Asian sufferers of MS may exhibit symptoms which are grouped into two different phenotypes: optic-spinal MS (OSMS) and conventional MS (CMS) (Ishizu et al. 2009). These phenotypes relate to clinical sites affected by MS rather than general clinical trends and will not be discussed further in this study.

2.2.2 Prevalence

The global median prevalence rate of MS is estimated to be around 30 per 100,000 ($30/10^5$) within a range of 5–80/ 10^5 (WHO and MSIF 2008). However, these are official country estimates and individual studies give much higher rates for many areas. For example, prevalence rates of 248/ 10^5 and 193/ 10^5 , have been reported for Saskatoon, Canada, and Orkney, Scotland, respectively (Rosati 2001). These rates represent nearly 1 in 400 and 1 in 500 of the local population. Geographical variations in prevalence rates of MS will be discussed further in Chapter 3. In the rest of this chapter, factors which affect prevalence will be discussed.

MS is thought to be multifactorial with both genetic and environmental factors playing a role. Many of the factors involved, and an indication of their interdependencies, are shown in Figure 2.4. It can be seen that a major factor is time, which can affect almost every aspect of MS through changes in the external physical and social environment, physical health, mental attitude, and lifestyle of an individual. Pre-onset symptoms may also affect other factors such as diet or smoking which could then feed back to influence the subsequent onset of MS. In addition, determination of the prevalence of MS can also change with time, depending not only on changes in case numbers, but also on changes in available health care resources, diagnostic criteria, population structures and analysis techniques.

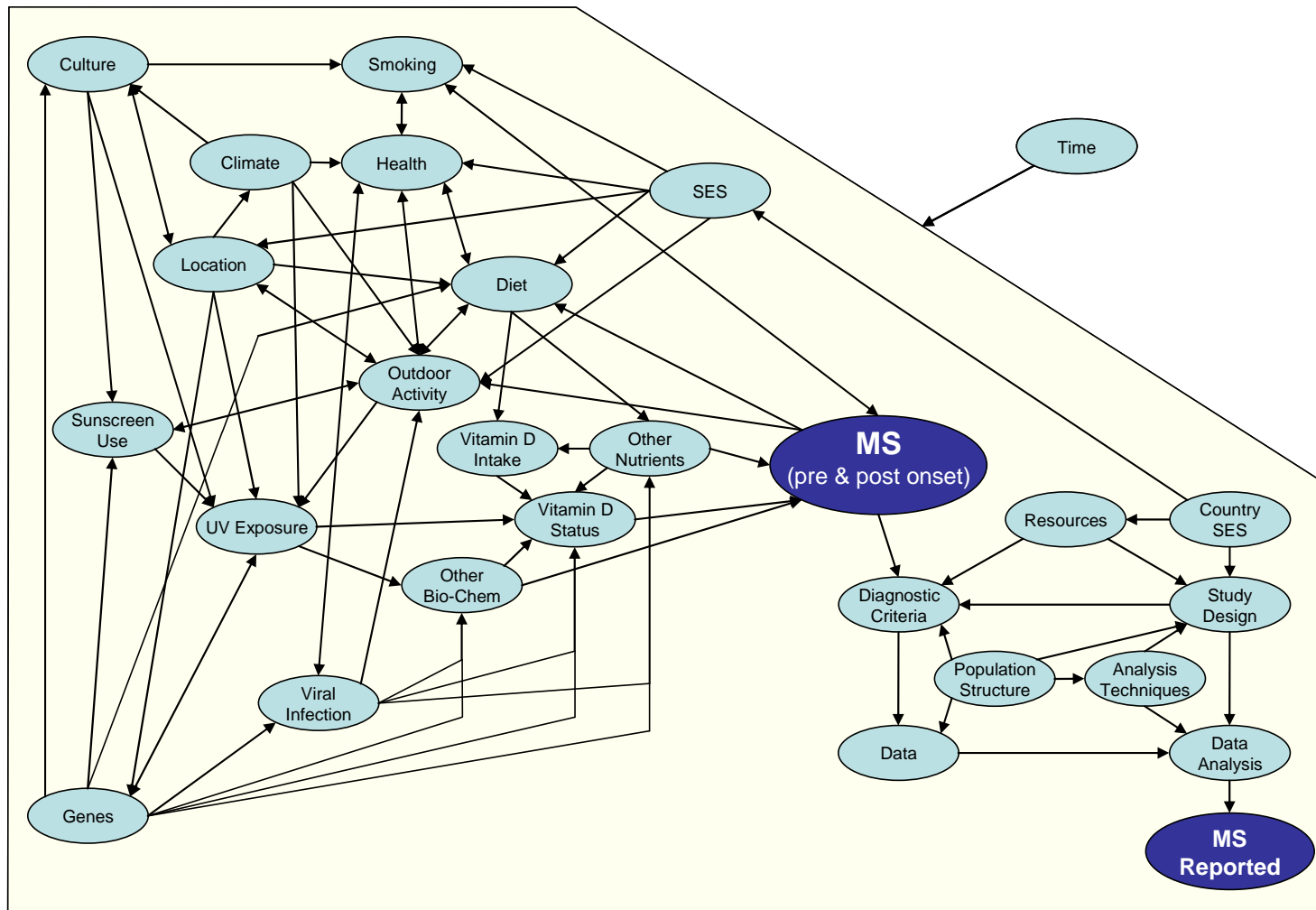


Figure 2.4 Factors affecting the prevalence of MS

Key factors affecting the prevalence of MS and their interdependencies. Arrows indicate the direction of effect.

Other factors thought to be involved in the prevalence of MS are generally grouped into genetic factors and environmental factors. Genetic factors tend to remain constant, other than UV damage to DNA which can occur, with most of the dynamic factors being in an individual's environment. Any study which looks at the prevalence of MS in conjunction with any single factor or group of covariables, should, therefore, be aware of the potential for confounding by other factors. Only if influences from these other factors can be isolated, or shown to be negligible can the covariables of concern then be considered as non-confounding covariables.

The strongest genetic factors include gender, ethnicity and family history of MS, while the strongest environmental factors include viral infection and latitude of residence. Some of these factors will now be outlined and then discussed further in section 2.2.3.

Gender: Women are found to be 2–3 times more likely to get MS than men, and this sex-ratio is increasing with time (Ebers and Goodin 2007; Ebers 2008; Greer and McCombe 2011; Koch-Henriksen and Sørensen 2010; Sellner et al. 2011). The ratio is higher for RR/SPMS than for PPMS (Sundström et al. 2001; Thompson et al. 1997).

Familial: There is a greater risk of two twins or siblings both having MS than two unrelated people in the general population (Dyment et al. 2004; Ebers and Goodin 2007; Ebers 2008; Goodin 2010; Islam et al. 2006; Kahana 2000). The risk is increased further if one or both parents also have MS.

Ethnicity: People of white, European ancestry are the most at risk. Cases within many indigenous populations are rare (Koch-Henriksen and Sørensen 2010; Pugliatti et al. 2002; Rosati 2001; Smestad et al. 2008).

Viral: Over 99% of sufferers test positive for antibodies to the asymptomatic Epstein–Barr virus (EBV), with the symptomatic infection of glandular fever, or infectious mononucleosis (IM), increasing the risk by a factor of 2–3 (Handel et al. 2010; Thacker et al. 2006).

Latitude: Prevalence rates tend to increase with distance from the equator, whether the northern or the southern hemisphere. This factor will be discussed briefly below and then looked at in more detail in the next Chapter on Geography.

There are, of course, exceptions to these trends, two of which can be immediately linked to ethnicity:

The Samis of northern Europe should have a high prevalence rate, being of European ancestry and living at high latitude. However, until recently there have been very few MS sufferers in the Sami population and the prevalence rate is much lower than that in other north European populations. The first recorded Sami case was in 1983 with one case who had a Sami father (Grønlie et al. 2000). The first cases where both parents were Sami were three cases recorded in 1993; three further cases had a single Sami parent. A later survey (Harbo et al. 2007) still only found 12 cases within the Sami, giving a prevalence rate of $5\text{--}30/10^5$, depending on the degree of Sami ancestry included in the calculation; this compared with $73\text{--}164/10^5$ in the non-Sami Norwegian population. Harbo et al note the low population numbers for the Sami, estimated to be up to 40,000 in Norway in 2004 (Statistics Norway 2010). They also note the uncertainties in both this population figure and the number of Sami with MS due to possible under-reporting. Even so, they go on to point out that the range of values reported is "significantly lower than the prevalence reported in other non-Sami Norwegian populations".

The Sardinians should have a relatively low prevalence rate, being one of the most southerly European populations in the northern hemisphere. However, their prevalence rate is one of the highest in the world at $144\text{--}152/10^5$ compared with $40\text{--}70/10^5$ for mainland Italy (Pugliatti et al. 2001b; Rosati 2001).

Other major exceptions to the general prevalence trends where strong ethnic groupings cannot provide a ready explanation are most likely due to lifestyle and environmental factors such as smoking, diet, and the amount of time spent in the sun:

Smoking: Higher prevalence rates are seen in smokers than in non-smokers (Hernán et al. 2005; Hernán et al. 2001; Riise et al. 2003). The sex-ratio of smokers is strongly correlated with the sex-ratio of the prevalence of MS (Palacios et al. 2011).

Diet: Regions where there is a high proportion of fish in the diet tend to have a lower prevalence rate than similar areas where less fish is consumed. Supplementing the diet with Vitamin D tends to reduce the risk of MS. (Ascherio and Munger 2007b; Kampman et al. 2007; Kampman and Brustad 2008; Kampman and Steffensen 2010; Pierrot-Deseilligny and Souberbielle 2010; Solomon and Whitham 2010).

Exposure to sun: Children and adolescents taking part in summer outdoor activities tend to exhibit lower prevalence rates than those who do not (Kampman et al. 2007;

Koch-Henriksen and Sørensen 2010). People born in early summer are more likely to get MS, those born in early winter are less likely (Salzer et al. 2010; Staples et al. 2010; Willer et al. 2004). It should be noted that reports of variations in the seasonal occurrence of relapses are mixed: McMichael and Hall (1997) report several studies which observed more relapses in winter than in summer, whereas Cross and Parks (2010) report studies which note more relapses in spring and summer. This highlights the dangers of drawing conclusions from just a few studies—observed seasonal variations may be subject to other confounding factors rather than an anticipated correlation with ambient exposure to sunlight. It also highlights the importance of considering possible seasonal variations when designing short term (6–9 month) studies which look at the pathology of MS. Putting these issues aside, however, there does seem to be support for a protective role for UV exposure and vitamin D intake in the MS pathology (Hanwell and Banwell 2011; Ponsonby et al. 2002; Ponsonby et al. 2005). (Sections 2.2.3 and 2.2.4 will look further at vitamin D and factors associated with vitamin D.)

Migration: Migration is a significant confounding factor for the latitude trend—people who migrate during childhood and adolescence tend to adopt the susceptibility of the place they move to; people who migrate in adulthood tend to keep the susceptibility of where they grew up (Cabre 2007; Cabre et al. 2005; McLeod et al. 2011; Wallin et al. 2009). However, as Marrie (2004) points out, "migrants tend not to be representative of their region of origin, typically being younger, healthier, and of higher socioeconomic status". As such, there may be other confounding factors to consider so any migrant population study should not be taken in isolation.

2.2.3 Factors Affecting Prevalence

Some of the factors mentioned above are now discussed further, though a conclusion by Koch-Henriksen and Sørensen (2010) should be kept in mind: "New insights into epistasis and epigenetics have ruled out the possibility of simple causative associations between genes or the environment and MS". So, although factors are listed here separately, many interact and should be considered as possible covariables.

2.2.3.1 Gender and Familial

Studies show that there is a significant sex ratio in the prevalence of RR/SPMS which is increasing with time (Sellner et al. 2011). The female to male (F:M) ratio is 2–3:1 for all MS cases, with the ratio for PPMS being lower at 0.6–2.1:1 (Sundström et al. 2001; Thompson et al. 1997). However, gender in MS is not a simple factor: both genetic and environmental elements are involved and many factors interact (Greer and McCombe 2011; Nicot 2009; Sellner et al. 2011).

A sex linked factor suggests that any genes or alleles which might be involved could lie on the X chromosome, and this could contribute to the sex ratio of the prevalence of MS. Although a few possible MS susceptibility loci have been identified on the X chromosome, none have yet been confirmed (Greer and McCombe 2011). However, section 2.2.3.4 describes an X chromosome linked retrovirus which appears to act with the Epstein–Barr virus to increase the susceptibility of individuals to MS. If this, or any other such X chromosome link, is confirmed, then it might explain some of the 2–3:1 sex ratio which is observed, though not explain it completely. Therefore, regardless of whether or not part of the sex ratio can be explained by an X chromosome linked factor, there must be at least one other factor involved which is due to environment.

Differences in environment between men and woman can be broadly separated into two categories: physiological and lifestyle. The main physiological difference is the presence (or absence) of various sex hormones. Evidence for hormonal influences comes mainly from post-onset prognosis: lower relapse rates during pregnancy and a worsening of symptoms during menstruation (Ascherio and Munger 2007b; Sospedra and Martin 2005; Tomassini and Pozzilli 2009). However, differences in genes and hormones during the early development and perinatal periods also cause gender differences within the CNS and the immune system, which may render women more susceptible to MS. Further information can be found in any of a number of recent reviews which give in depth details of gender, hormones, cytokines and potential molecular mechanisms in MS (Akdis et al. 2011; Eikelenboom et al. 2009; El-Etr et al. 2011; Greer and McCombe 2011; Kuhlmann et al. 2009; Moldovan et al. 2008; Nicot 2009).

Aspects of lifestyle which can differ between men and women, at a population level, are largely determined by culture. Differences may include: the amount of time spent

outdoors; whole body clothing; diet; smoking; obesity; and assessment and health care bias (Sellner et al. 2011). The first two of these, the amount of time spent outdoors and whole body clothing, could contribute to the sex ratio as discussed in section 2.2.3.6, below, as may diet, discussed in 2.2.3.5. Obesity increases the risk of MS. Whether obesity contributes to the sex ratio of MS depends partly on the relative risk between women and men, and also the numbers of obese women compared to men. Although the numbers would vary for different countries and cultures, a survey in New Zealand in 1997, found a slightly higher number of males than females categorised as overweight or obese (Rockell et al. 2006) so the relative risks due to obesity are likely to be more important than the relative numbers in this case. (Obesity is discussed further in section 2.2.3.5 on diet.) Assessment and health care bias may vary by country and culture. However, the trend towards more women seeking medical advice than men, may contribute to any change in sex ratio rather than the ratio itself. The last lifestyle factor listed here is smoking which is covered in more detail in section 2.2.3.3. Although smoking can explain the change in sex ratio with time, it cannot explain the underlying sex ratio.

The degree to which risk of a disease can be considered genetic as opposed to environmental can be established by studying the prevalence of the disease within families: if there is a genetic element then the degree of shared genetic material should be reflected in concordance rates. So if the concordance rates for MS in related individuals are greater than those for the general population then there must be a genetic element to the prevalence of MS. Similarly, lower concordance rates for related individuals raised in different environments will point to environmental factors.

The amount of shared genetic material varies between monozygotic (MZ) and dizygotic (DZ) twins, siblings, half siblings, parents/children and more distant relatives, and the concordance rates for MS vary accordingly (Dyment et al. 2004; Ebers 2008). Table 2.1 gives a selection of concordance rates from Canadian studies (Ebers and Goodin 2007; Ebers 2008; Willer et al. 2003). The rates observed are consistent with a strong genetic link to the prevalence of MS, yet the 34% concordance for MZ twins (compared to a full 100%) implies that environment factors are also important. It is also noticeable that the rates for maternal and paternal half siblings differ, as do the rates for female and male MZ twins. This suggests that gender is a factor, with a strong maternal

Pairing	Prevalence per 1000	Genes Shared
General population	1	0%
Adopted siblings	1	0%
Cousins	7	12.5%
Paternal half siblings	13	25%
Maternal half siblings	24	25%
Full siblings	35	50%
Dizygotic twins	55	50%
HLA-identical siblings	80	50%
Offspring of consanguineous pair	90	56.25%
Child/Parent	197	50%
Offspring of conjugal MS pair	200	50%
Monozygotic twins	270	100%
Male monozygotic twins	65	100%
Female monozygotic twins	340	100%

Table 2.1 Selected prevalence rates in relatives of MS probands

Concordance rates for prevalence of MS from the Canadian studies reported by Willer et al (2003), Ebers and Goodin (2007), Ebers (2008) for the selected pairs of relatives.

'parent-of-origin effect'. Previous studies in Denmark (Nielsen et al. 2005) and North America (Islam et al. 2006) also found a strong genetic factor in concordance rates, though the maternal parent-of-origin effect was not apparent in the Danish study and not as strong in the North American study. Although these studies all confirm a genetic factor, they do not identify, or attempt to identify, any specific gene or genes.

Using concordance rate data from Willer et al (2003) and data for the frequency of the HLA DRB1*1501 allele (see section 2.2.3.2), Goodin (2010) used a model for susceptibility in MS to predict that a relatively small number of possible susceptibility loci (11–18 out of a possible 50–200) combine to give the overall susceptibility to MS. (However, as Goodin points out, the model cannot distinguish between true susceptibility-loci and disease-modifying loci.) In addition, the model predicts that less than 2.2% of the general population is genetically susceptible to MS. This would put an upper limit on the prevalence of MS, and it also suggests that only a fraction of those who are genetically susceptible to MS are diagnosed as having MS; other factors must, therefore, be involved.

2.2.3.2 Ethnicity and Other Genetic Factors

The genetic European family tree indicates why the Sami and the Sardinian populations are so notable, and so useful in MS research. As shown in Figure 2.5, the

Sami population is the oldest population in Europe, with the Sardinian branch not far behind. The Sami population migrated from south western Europe after the last great ice age, then settled and remained genetically isolated in the north of Scandinavia. Sardinia was one of the first areas to be settled in the Mediterranean region and, despite frequent invasions by other races, the Sardinian population tended to avoid interbreeding, and has thus retained its own particular gene pool (Pugliatti et al. 2002). Other European ethnicities all diverged after the Sardinian branch. Given the two extremes in the MS prevalence rates exhibited by these two populations, an understanding of how their genomes, and those of the other ethnicities, differ should help to pinpoint which genes to focus on in investigations into the genetic causes of MS (Grønlie et al. 2000; Harbo et al. 2007; Pugliatti et al. 2001a; Pugliatti et al. 2001b; Stefano et al. 2003). Also of interest is that the Scottish and Irish ethnic branches diverged before many of the other European branches, with the English branch diverging much later. Since prevalence rates in Scotland and Ireland are 30–60% higher than in England, a difference which is unlikely to be explained by latitude alone

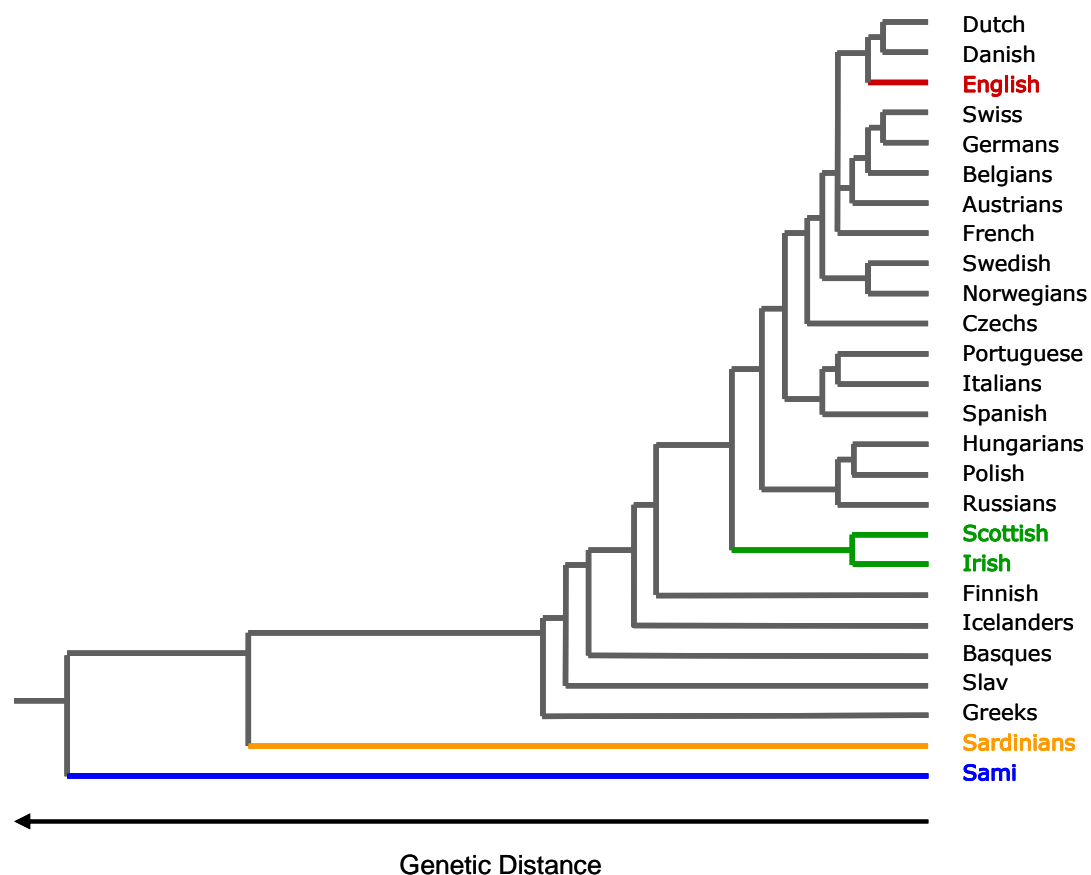


Figure 2.5 The European family tree

After Figure 4 of Ebers (2008). Several ethnic branches are highlighted. Blue: Sami; Orange: Sardinian; Red: English; Green: Scottish and Irish.

(Pugliatti et al. 2006; Robertson and Compston 1995), there may be a genetic factor involved, which may be related to specific genes (see below) or to skin type.

Many ethnicities have evolved a skin type suitable for the latitudes at which they evolved—most equatorial ethnicities have dark skins to protect against excess UV whereas most of the ethnicities which evolved nearer the polar regions have fair skin (Holick et al. 2007). With reference to the Scottish, Irish and English ethnicities, it is interesting to note that the Scottish and Irish have the fairer type I skin whereas the English have type II skin (McKenzie et al. 2009; Zhang and Naughton 2010). Skin type and pigmentation will be mentioned further in section 2.2.3.6.

Genetic based research has changed radically over the last decade or two, mainly due to the Human Genome Project combined with major advances in computational analysis techniques. With these advances also came a need for a new way of thinking about genetic research, especially when looking at complex autoimmune associated diseases, such as type I diabetes, rheumatoid arthritis and MS (Baranzini 2011; Brookes 1999; Gourraud et al. 2012; Hoffjan and Akkad 2010; Rioux and Abbas 2005; Wandstrat and Wakeland 2001). It is interesting to note that many genes are now believed to be associated with many autoimmune diseases, as illustrated in Figure 1 of Baranzini (2011). This highlights the commonality between many of the pathogenesises of the various diseases (Baranzini 2011; Rioux and Abbas 2005; Wandstrat and Wakeland 2001; Zhernakova et al. 2009). An in depth review of what is now known in conjunction with MS is well beyond the scope of this dissertation. However, a broad outline of certain aspects will now be given.

Many genetic factors which affect the prevalence of MS act through the presence or absence of particular alleles or haplotypes (combinations of alleles) in the major histocompatibility complex (MHC) which moderate elements of the immune system. In particular, a number of human leukocyte antigen (HLA) alleles have been found to be positively associated with MS, and others have been found to be negatively associated. Extensive reviews on aspects of MS which include discussions of HLA, MHC and non-MHC, alleles and haplotypes, have been published by many authors (Baranzini 2011; Boppa et al. 2011; Compston and Coles 2008; Gourraud et al. 2012; Hoffjan and Akkad 2010; Kakalacheva and Lünemann 2011; Kakalacheva et al. 2011; Lincoln and Cook 2009; Ramagopalan et al. 2010; Shapira et al. 2010; Sotgiu et al. 2004; Zuvich et

al. 2009). Although a number of alleles and haplotypes associated with risk of MS have now been identified, understanding how these affect the pathogenesis of MS is still far from clear. However, the allele HLA DRB1*1501 appears to be most highly associated with risk of MS and could be involved as follows. Through a series of intermediate receptors, vitamin D is thought to bind to HLA DRB1*1501 (Handunnetthi et al. 2010). Low vitamin D, and low binding and regulation of HLA DRB1*1501 may affect T-cell mediated auto immune responses, and impaired T-cell activity is believed to be important in MS pathogenesis (Chastain et al. 2011; Gandhi et al. 2010; Khoury et al. 2000; Link et al. 1992; Ramagopalan et al. 2009; Sato et al. 2011; Skulina et al. 2004; Tumani et al. 2009; Wucherpfennig and Sethi 2011). However, only around half of MS sufferers have the HLA DRB1*1501 allele, and its presence does not necessarily mean that an individual is susceptible to MS (Goodin 2009). Therefore, it is probable that this allele is just one of many alleles associated with susceptibility to MS and that MS is multigenic with the overall genetic susceptibility determined by an as yet unknown combination of alleles (Baranzini 2011; Goodin 2010; IMSGC and WTCCC 2011; Sotgiu et al. 2004). In an Oslo study, Spurkland et al (1997) found that 95% of cases and 98% of controls who had either the HLA DRB1*1501 or DQB1*0602 alleles carried these as the HLA DRB1*1501-DQA1*0102-DQB1*0602 haplotype. Although that study was directed at the few cases without this haplotype, it serves to highlight one of the difficulties in establishing susceptibility due to individual alleles. In addition, Harbo et al (2007) found that the general Sami population has a much lower frequency of this specific HLA DRB1*1501-DQA1*0102-DQB1*0602 haplotype than the general Norwegian population. The relative lack of this haplotype in the Sami, could, therefore, be a contributory factor in the low prevalence of MS in the Sami. In contrast, the Scottish and Northern Irish populations appear to have a higher frequency of the HLA DRB1*1501 allele than the population of southern England, which may help to explain the very high prevalence rates experienced in Scotland (Lonergan et al. 2011; McGuigan et al. 2004; Swingler and Compston 1986).

Recent advances in technology and analysis techniques have enabled many more studies to assess correlations between the prevalence of MS and genetic risk factors (Fernández et al. 2008; Harbo et al. 2007; Lonergan et al. 2011; Nischwitz et al. 2011; Schmidt et al. 2007; Simpson et al. 2011), including several as genome-wide association studies (GWAS) which study single nucleotide polymorphisms (SNPs)

(IMSGC and WTCCC 2011; Kakalacheva and Lünemann 2011; Patsopoulos et al. 2011). (SNPs are alleles with variations at a single nucleotide site, and where the least frequent allele has an abundance of at least 1% (Brookes 1999).) GWAS can assist in identifying correlations between single and multiple genetic variants and complex diseases such as MS within very large data sets of thousands, or tens of thousands, of cases and controls, and hundreds of thousands, or millions, of SNPs. However, there is always a degree of uncertainty since GWAS perform many tests for significance simultaneously for which a single threshold of significance must be set, and there is always the question of what level this should be at. In addition, most GWAS include data from different populations, so genetic heterogeneity must be allowed for to avoid confounding. Despite these considerations, studies have been successful at identifying many more loci associated with the susceptibility of MS. The most significant loci are still those in the HLA region, however, many more have been discovered that are non-HLA and which are of much more modest significance (Gourraud et al. 2012).

Having identified various susceptibility loci, there is then still the task of determining how each gene is involved in the susceptibility of MS. It is now believed that many genes, each with modest affect, combine to give an overall susceptibility, as illustrated for T-cell function in Figure 2.6 (after Figure 2.A of Gourraud et al (2012)) where each function category shown involves at least four (and as many as 15) different genes, and the links represent commonality of at least two genes. With such a network of links, the genetic susceptibility for T-cell function is not straightforward.

As mentioned in section 2.2.2, above, and section 2.2.3.5, below, low vitamin D status appears to increase the risk of MS, and vitamin D supplementation can alleviate some of the symptoms, possibly by altering T-cell homeostatis (Correale et al. 2009).

Although the HLA DRB1*1501 allele has been highlighted specifically for risk associated with MS, many other genes are associated with risk of vitamin D insufficiency, as confirmed by the GWAS of Wang et al (2010). Given that it is probable that there are numerous genes and alleles involved in determining the susceptibility of an individual to MS, then it may well be that studies aimed at establishing the genes and alleles involved in MS also need to look at those involved with other associated factors (Sospedra and Martin 2005).

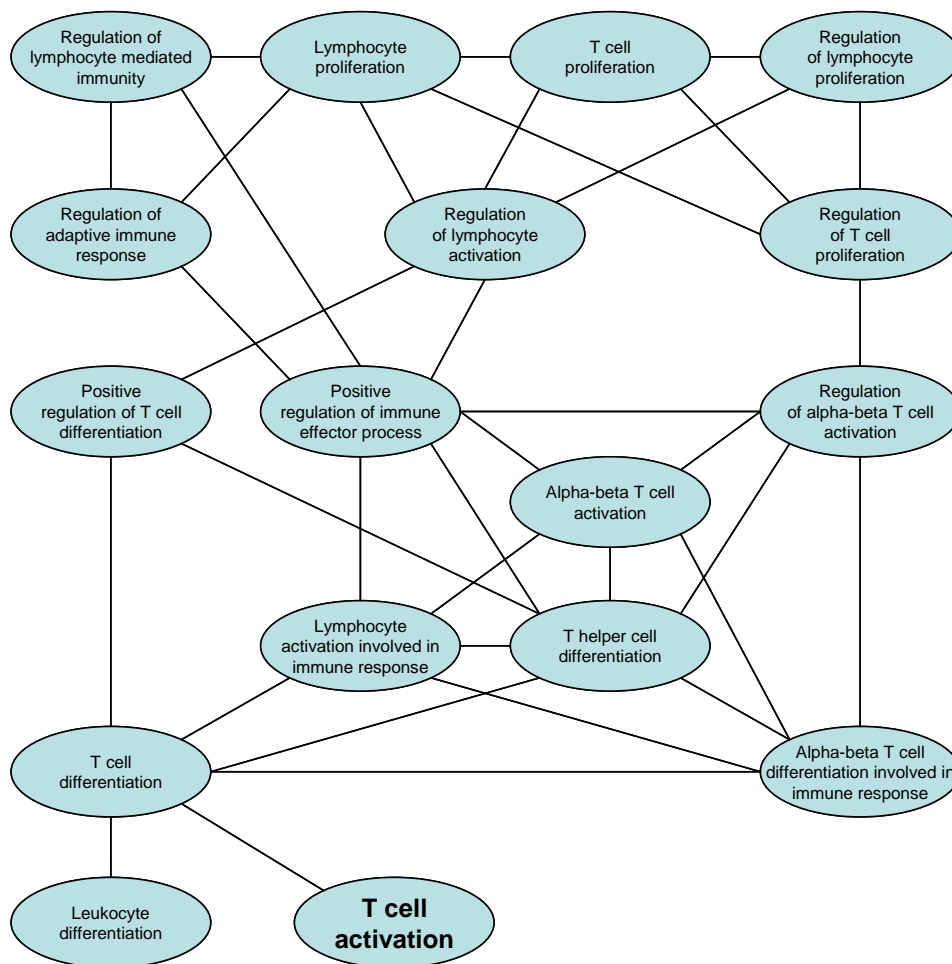


Figure 2.6 T-cell biological pathways associated with a susceptibility to MS
 Commonality of genes associated with T-cell function and MS, after Figure 2.A of Gourraud et al (2012). Categories shown each have at least four genes; links indicate at least two shared genes.

One such group of candidates are tumor necrosis factor (TNF)- α alleles, two of which (TNF₋₃₇₆ and TNF₋₂₃₈) are associated with the high prevalence of MS in Sardinia (Sotgiu et al. 2004). These (along with TNF₋₃₀₈) are usually associated with a susceptibility to malaria, and, although both have a low frequency throughout most of the world, they are particularly elevated in the Sardinian population where malaria was endemic until the end of the 1940s. TNF- α alleles express proinflammatory cytokines which have been found in active MS lesions (Sospedra and Martin 2005) and the cerebrospinal fluid of MS patients (Drulović et al. 1997). TNF- α has also been associated with an increased risk of MS in Omsk (Khanokh et al. 2011).

Other susceptibility loci which have been determined by various studies include genes which express for various interleukins (ILs), TNF- β and interferon (IFN)- γ . Differences in the presence of certain interleukins and TNF- β have been found in blood cells from

SPMS cases when compared with RRMS cases (Rovaris et al. 2006), and other interleukins have been found to disrupt the human blood–brain barrier (Compston and Coles 2008). Yang et al (2011) provide an in depth description of the important role of the interleukins IL-12 and IL-23 in T-cell function as it relates to the pathogenesis of MS. IFN- γ also plays a part, though whereas it was once thought that IFN- γ had a key role in MS (Traugott and Lebon 1988), it would now appear that IL-17 is more important (Compston and Coles 2008). However, IFN- β remains important in the treatment of MS (Boppana et al. 2011; Hauser and Oksenberg 2006; Maghzi et al. 2011a; Rauchway 2011). Cytokines such as IL-2, IFN- γ and TNF- β are also produced by T-cells after reactivation by the Epstein–Barr virus (see section 2.2.3.4). These may then be involved in myelin destruction (Pender 2011). Further details and examples of the roles of cytokines and related factors in MS and the CNS can be found in many other sources (Bar-Or et al. 1999; Bettini and Vignali 2009; Diveu et al. 2008; Imitola et al. 2005; Laroche et al. 2011; Matusevicius et al. 1996; Merson et al. 2010; Muñoz-Fernández and Fresno 1998; Navikas et al. 1996; Petermann and Korn 2011; Sosa and Forsthuber 2011; Wu and Alvarez 2011). For further details of the genetics and loci involved in susceptibility to MS, the reader is directed to the cited literature.

2.2.3.3 Smoking

There are a number of recent reviews and meta-analyses which summarise and discuss the evidence for a connection between smoking and an increased risk of MS (Handel et al. 2011; Hawkes 2007; Jafari and Hintzen 2011; Palacios and Simon 2012; Wingerchuk 2012). An overview is given here.

Results from two surveys of women nurses in the US in 1976 and 1989 found that smoking had a relative incidence rate of 1.6 compared to women who had never smoked, and past smokers had a relative rate of 1.2. The rates obtained also increased with cumulative exposure to smoking (Hernán et al. 2001). Similar results were obtained for a study of both men and women in the UK in 2000: ever having smoked gave an odds ratio of 1.3 (Hernán et al. 2005). This risk was the same for PPMS and RRMS, however, smoking then gave a hazard ratio of 3.6 for RRMS progressing to SPMS. In Norway in 1997, ever having smoked gave a rate ratio of 1.8 for MS (Riise et al. 2003). More recently, Palacios et al (2011) looked at smoking and the sex ratio for MS. As described in section 2.2.3.1, the F:M sex ratio of MS prevalence has been

increasing with time. In addition, over the last century in developed countries, there has been a general increase in the number of women who smoke and a general decline in the number of men who smoke, leading to an increase in the F:M sex ratio of smokers. Looking at data from Canada and Denmark, the authors determine that smoking can potentially explain 90% and 94%, respectively, of the increase in the sex ratio of MS. It remains unknown, however, whether other factors which may also be connected with gender are acting alongside the effect of smoking, thus confounding the results. Such factors could include changes to diet and exposure to sunlight, which are discussed later in sections 2.2.3.5 and 2.2.3.6. It would also be interesting to see if the increase in smoking in women is linked to the increase in the sex ratio of RR/SPMS (as opposed to all MS), since the rise in the overall sex ratio appears to be driven by this phenotype.

The studies described above do not, however, explain why smoking has such an impact on the prevalence of MS (or explain the underlying sex ratio). It is apparent, though, that the risk of MS is associated with inhalation of the tobacco smoke, and not, for example, due to nicotine in the tobacco since snuff use does not appear to increase the risk of MS (Hedström et al. 2009). Rather, prolonged use of snuff decreases the risk of MS, possibly due to nicotine acting as a neuroprotective agent. In addition, there is evidence that passive smoking amongst non-smokers can increase the risk of MS (Hedström et al. 2011a; Mikaeloff et al. 2007). If this is the case then passive smoking could be confounding the results in studies which only consider 'never' and 'ever' having smoked.

Although the precise biological pathways connecting smoking and MS are not currently known, tobacco smoke, or specific chemicals within tobacco smoke, are known to be associated with processes of other diseases (Hawkes 2007; Hedström et al. 2009; Hedström et al. 2011b; Jafari and Hintzen 2011; Lincoln and Cook 2009; Mikaeloff et al. 2007). Processes which are also associated with MS include: abnormalities in T-cell function; increased blood–brain barrier permeability; demyelination caused by thiocyanate and cyanide intoxication; axonal degeneration caused by free radicals or nitric oxide exposure; increased respiratory infections; deficiency of IFN- γ producing cells; impairment of humoral and cell-mediated immunity; immunomodulation or pro-inflammatory influence on the immune system; increased cellular apoptosis; and

modulation of various proinflammatory cytokines, chemokines and cell surface receptors. It is possible, therefore, that tobacco smoke is a facilitator in the role of some or many other factors (Pierrot-Deseilligny and Souberbielle 2010). Further to this, a study by Simon et al (2010) has demonstrated that smoking can enhance an association between the HLA DRB1*1501 allele and the Epstein–Barr virus, discussed in section 2.2.3.4. Similarly, Hedström et al (2011b) have provided evidence that smoking enhances the association between MS and the positively associated HLA DRB1*15 and the negatively associated HLA A*02 alleles. However, as Sawcer and Hellenthal (2011) point out in their commentary, reliable analyses of gene-gene and gene-environment interactions are not easy to perform, and whether or not confounding factors such as selection bias have been adequately controlled is not easy to determine.

2.2.3.4 Epstein–Barr Virus

The Epstein–Barr virus (EBV) is an infection which most people get in childhood—90% of the population test positive for antibodies to the virus, though most do not express any clinical symptoms. Infection in late adolescence or adulthood, however, usually manifests itself as the symptomatic glandular fever, or infectious mononucleosis (IM). When it comes to MS, EBV is an enigma (Ludwin and Jacobson 2011; Pohl 2009). As mentioned in section 2.2.2, over 99% of MS sufferers test positive for antibodies to the virus and IM infection increases the risk. In addition, not having EBV antibodies appears to reduce the risk of MS by a factor of 10 (Ascherio and Munger 2007a). Is it the EBV infection itself which leaves people susceptible to MS or is there some other precursor which makes people susceptible to both EBV and MS? Although EBV was long thought of as being a precursor to MS, some researchers are now asking whether this is truly the case, or whether EBV just happens to be unfortunate enough to manifest itself sooner than MS, and so is simply in the "wrong place" at the "wrong time" (Maghzi et al. 2011b) and "guilty by association" (Lünemann and Münz 2009). It should also be noted that the clinical features of MS are only manifested if the underlying neurological changes pass a clinical threshold (Ramagopalan et al. 2010). Sub-clinical neurological damage could be occurring long before the first clinically observable attack so the timeline between EBV and onset of MS may not be as straightforward as it appears.

It is also interesting to note, however, that EBV antigens may bind to vitamin D receptors or other receptors associated with MS, by molecular mimicry or other means, and block activation of its target genes. The EBV antigens may also directly affect T-cell activity (described in section 2.2.3.2). Each of these possible mechanisms could lead to an increased susceptibility to MS (Brennan et al. 2010; Christensen 2006; Disanto et al. 2011; Kakalacheva and Lünemann 2011; Kakalacheva et al. 2011; Lucas et al. 2011a; Pender 2011; Pender et al. 2009; Yenamandra et al. 2010). In addition, it appears that the presence of the HLA DRB1*1501 allele increases the ability of the immune system to react to the EBV virus, thus increasing the risk of MS even further (Sundström et al. 2009). As mentioned above, over 99% of MS sufferers test positive for EBV antibodies. Nociti et al (2010) established that almost all cases in their study of individuals with either of two other neurological disorders, chronic inflammatory demyelinating polyradiculoneuropathy (CIDP, the equivalent of MS for the peripheral nervous system) and amyotrophic lateral sclerosis (ALS, a degenerative disorder of aging nervous systems), also had EBV antibodies in blood serum, providing evidence of previous EBV infection. However, the levels of EBV antibodies were significantly higher in the MS cases compared to those in the CIDP and ALS cases. In addition, EBV antibodies were found in the cerebrospinal fluid of the MS cases, but not in that of the CIDP and ALS cases. After allowing for potential confounding by age (the MS cases were significantly younger than the CIDP and ALS cases) the authors suggest that the increased levels of EBV antibodies in MS cases provide further evidence of an involvement of EBV infection in MS. It should also be noted that antigens associated with different strains of EBV may occur at different frequencies in MS cases compared with controls (Brennan et al. 2010), further complicating any MS–EBV relationship.

Recently, a new family of human endogenous retroviruses (HERVs), HERV-W, has been discovered which can be reactivated by herpes viruses such as EBV. See Perron et al (2009) and Perron and Lang (2010) for in depth reviews; a brief summary now follows. Endogenous retroviruses make up about 8% of the human genome and, being both genes and viruses, produce antigens which may or may not be tolerated by the immune system. In addition, retroviral sequences tend to occur as multiple copies, possibly in multiple genes, so reactivation is likely to cause an immune cascade. In the case of HERV-W, which has also been termed the MS-associated retrovirus (MSRV), expression of immunopathogenic proteins in susceptible individuals can lead to MS

lesions and disease progression. HERV-W could thus be a co-factor with EBV in the pathogenesis of MS. Just as notable is that the most likely location for copies of the MSR sequence is on the X chromosome, which might contribute to the sex ratio seen in the prevalence of MS, as discussed in section 2.2.3.1.

The possible introduction of an infectious agent to the Faroe Islands by British troops during World War II has also been proposed as a precursor to the first evidence of MS in the Faroe Islands (Kurtzke 2000; Kurtzke and Heltberg 2001). However, it is still unclear whether this is the case (Binzer et al. 2010, 2011; Wallin and Kurtzke 2011a), and if it is, what the infectious agent might have been (Carolei and Sacco 2011; Wallin and Kurtzke 2011b). If EBV was involved then it could be that the British introduced a new, MS-related strain, since EBV was almost certainly present in the Faroe Islands before the British troops arrived (Ascherio and Munger 2007a).

2.2.3.5 Diet

There are many aspects to diet which affect human health, though the over-riding one of note when considering the prevalence of MS is the intake of vitamin D. (Dietary factors such as meat preservation methods, which are thought to be linked to a risk of MS (Lauer 2007), will not be discussed here.) A low vitamin D status appears to be a risk factor for MS, and may also influence disease progression (Ascherio and Munger 2007b; Ascherio et al. 2010; Hanwell and Banwell 2011; Kampman and Brustad 2008; Kampman and Steffensen 2010). What is 'low', however, has yet to be established, and blood serum levels which define hypovitaminosis D, vitamin D insufficiency or vitamin D deficiency are still open to debate (Rovner and O'Brien 2008), as discussed in section 2.2.4.2, below.

Vitamin D is only found naturally in significant quantities in a few foods, most notable of which are oily fish, marine mammals and a few types of mushrooms (Deutch et al. 2007; Holick 2006; Huotari and Herzig 2008; Juzeniene et al. 2011; Kuhnlein et al. 2006; O'Mahony et al. 2011; Phillips et al. 2011; Zhang and Naughton 2010). Some authors have referred to reindeer meat as supplying vitamin D to natives of Lapland (Gillie 2006; Goodin 2009), an assumption probably made by other authors at some point in the past and then propagated through the literature for want of contradictory evidence. (The diet of the reindeer includes lichen which is rich in vitamin D, and the stomach contents of herbivores were sometimes used in traditional Arctic foods).

However, recent studies have shown that meat (including liver and tallow) from semi-domesticated reindeer, grazing in traditional pastures, contains no detectable vitamin D (Hassan et al. 2012). Similarly, meat from the related caribou of northern Canada has also been shown to contain little or no detectable vitamin D (Kuhnlein et al. 2006).

Fish, marine mammals and reindeer (or caribou) form the basis of the traditional diet of many northern, and especially Arctic, native races (Deutch et al. 2007; Johnson et al. 2009; Nilsson et al. 2011). In addition, oils collected from the sea harvest (for example, cod liver oil in Norway) provide additional nutritional supplements (Brustad et al. 2004a; Brustad et al. 2004b; Huotari and Herzig 2008; Johnson et al. 2009). In Norway, almost half of the dietary vitamin D intake is due to supplements such as cod liver oil rather than through the direct consumption of fish (Jorde and Børnaa 2000). So although a high fish consumption is often used to explain why the people of northern Scandinavia show a lower prevalence of MS than might be expected from their high latitude location, it is possible that, when considering diet and MS, the Scandinavian tradition of supplementing vitamin D through cod liver oil is just as important as the consumption of fish, if not more so.

It should be noted that the presence or absence of traditional foods in the diet has a marked affect on the vitamin D intake of northern native populations, with vitamin D levels in the body often being several times higher after consuming traditional foods (Brustad et al. 2003). This is especially important during winter months when the reduction in available sunlight affects the body's own capacity for producing its own vitamin D (see section 2.2.4, below). Changes in lifestyles within various native populations are now having an affect on diet and these traditional foods are being displaced from the diet and being replaced by commercial, so called 'market food', with an associated decline in vitamin D intake (Brox et al. 2003; Kuhnlein and Receveur 2007). Despite this trend, however, Norwegians, for example, are still some of the top consumers of fish in Europe (Welch et al. 2002).

In addition to foods which are naturally rich in vitamin D, and supplements, whether traditional or commercially produced, another source of dietary vitamin D comes in the form of foods which have been artificially fortified. The fortification of foods with vitamin D and/or other micronutrients occurs mainly in industrialised countries and is often a contentious issue with much debate (de Lourdes Samaniego-Vaesken et al.

2012; Fletcher et al. 2004; Gillie 2010; Tylavsky et al. 2006). Whether or not foods are fortified, and the actual levels of fortification, vary considerably from country to country (de Lourdes Samaniego-Vaesken et al. 2012; Lawrence et al. 2009; Ovesen et al. 2003). Some countries (for example, Australia, Canada, the United Kingdom, and the United States) have a mixture of mandatory and voluntary fortification, others (for example, Denmark, Sweden, Finland and New Zealand) allow fortification in some foods, whilst others (for example Japan) have little or no food fortification. In addition, some countries (for example, Denmark) allow voluntary fortification but only under strict regulation, and fortified foods must be authorised before they can be marketed (DKMFLF 2011, 2012a, 2012b). The boundaries of a country, however, cannot be relied upon to determine whether certain food items are or are not fortified. For example, Australia has mandatory fortification of margarine and similar spreads with vitamin D, and, although New Zealand does not have mandatory fortification, most of the spreads consumed in New Zealand are imported from Australia, and are, thus, fortified. In the context of this study, it should be noted that vitamin D fortification of margarine and spreads in New Zealand has only been allowed since 1996 (Rockell et al. 2008). This highlights the situation that, in addition to there being differences between countries at any given time, the status of food fortification in any given country changes with time. Note also that there are two forms of vitamin D: vitamin D₃ is found in animal sources whereas vitamin D₂ comes from plant sources. Supplements and fortified foods can contain either, though the more potent vitamin D₃ (Heaney et al. 2011) is now usually preferred over the less potent vitamin D₂.

Another potential link of diet to MS is obesity – there is an increased risk of MS in overweight and obese individuals. Since vitamin D is stored in body fat, this link with obesity appears to be related to a low vitamin D status, though the full relationship between weight and vitamin D status is still less than certain (Arunabh et al. 2003; Blum et al. 2008; Heaney et al. 2009; Lagunova et al. 2009; Sellner et al. 2011). One widely accepted explanation is that an excess of body fat causes more vitamin D to be stored, thus depleting the levels of vitamin D circulating in the blood (Alshishtawy 2012; Blum et al. 2008; Brustad et al. 2003; Ebers and Goodin 2007; Ponsonby et al. 2010; Rockell et al. 2008; Smith et al. 2009; Thompson 2007; Zhang and Naughton 2010). However, vitamin D is not sequestered in body fat; rather, concentrations of vitamin D across the body as a whole are in balance. In addition, it has also been

hypothesised that obesity is itself caused by a low vitamin D status invoking a 'winter response' to prepare the body for winter conditions by storing fat both for insulation and energy (Foss 2009). This hypothesis would not necessarily preclude some element of vitamin D storage to also be involved in determining vitamin D status, however, extra vitamin D storage could in turn invoke a feedback loop which may be difficult to reverse until the vitamin D status returns to a sufficient level. Vitamin D status and the effect of winter is discussed in section 2.2.4.2.

2.2.3.6 Latitude and Sunlight

Easier to comprehend as a factor for MS prevalence is latitude, though that doesn't mean that the underlying aetiology is any closer to being understood. Bearing in mind other lifestyle factors, and combined with seasonal evidence, the dependence of the MS prevalence rate on latitude points to some dependence on sunlight, since the sunlight incident on the earth varies with season and latitude. For now it is sufficient to note that higher latitudes see less sunlight than lower latitudes, and winter months see less sunlight than summer months. This appears to be the inverse to the general behaviour of the MS prevalence rates and any protective role of increased UV exposure in the MS pathology. It is logical to assume, therefore, that there might be some correlation between a lack of sunlight and the prevalence of MS. In other words, the lack of sunlight appears to play a role in both the aetiology and the pathogenesis of MS. However, if there is a correlation between lack of sunlight and prevalence rates, variations in prevalence rates with month of birth still need to be explained.

As mentioned in section 2.2.2, people born in early summer are more likely to get MS and those born in early winter are less likely (Salzer et al. 2010; Staples et al. 2010; Willer et al. 2004). This would only be consistent with the above suggestion that a lack of sunlight is important in the aetiology of MS if the critical period of exposure to sunlight was not the time of birth but instead was some other period relative to time of birth. The time of least availability of sunlight is around mid winter, so an early summer maximum in prevalence rates would be consistent with a negative correlation between MS prevalence rates and either 4–5 months prior to birth (4–5 months into pregnancy) or an age of 7–8 months. Strictly speaking, the age of 7–8 months should be considered as 7–8 months plus any number of years. However, natural variations in how individual children develop suggest that any month effect is likely to dissipate

with age. Therefore, it is more likely that the critical period for any month effect is during pregnancy rather than after birth. In addition, the critical period of exposure to sunlight may be shifted relative to the time of least availability of sunlight depending upon whether the effect is a cumulative effect or the effect of some threshold. Comparisons of data from different latitudes with different lengths of summer and winter seasons should help to clarify this effect. The timing of exposure to UV is discussed in more detail in section 2.2.5.

The most likely link between UV and MS is vitamin D. The link between vitamin D and MS was discussed above in relation to diet; the link with UV will be discussed in depth in the next section. Indeed, the link between UV, vitamin D and MS has been so strong that other candidates for the role of linking UV to MS are only now being looked into in earnest in conjunction with MS aetiology and pathology (Bains 2010; Behan and Chaudhuri 2010; Behan 2010; Krone and Grange 2011; Mehta 2010; Perron et al. 2009). One such candidate is urocanic acid (UCA) which has long been associated with immunosuppression (though not specifically with MS) through T-cell activity, TNF, IFN and various other cytokines (Artuković et al. 2010; Hug et al. 1998; Mohammad et al. 1999; Noonan and De Fabo 1992). However, it appears to have a mixed role with different chemical pathways leading to competing pathogenesises. Also, whereas the different vitamin D pathogenesises are beneficial, those for urocanic acid can tip the immune system balance either way. Where the UV driven balance ends up depends on many factors which are still being investigated (Gibbs and Norval 2011; Halliday 2010; Hanwell and Banwell 2011; Lucas et al. 2011b; McKenzie et al. 2009).

Another contender for a link between latitude and MS is temperature. Moynihan and Moore (2010) hypothesise that heat acclimatization in children affects renal function. Through actions of uric acid, this then affects the blood–brain barrier and subsequent demyelination. Although this hypothesis will not be discussed further here, it serves as a reminder that there are potential risk factors for MS other than UV radiation which are related to latitude.

Further discussion of sunlight will be given in Chapter 6. The next section provides an overview of vitamin D, since it has been highlighted as being involved in the risk of MS.

2.2.4 Vitamin D—the Sunshine Vitamin

Vitamin D intake through diet was discussed in section 2.2.3.5. This section will give an overview of current understanding of other aspects of the biochemistry of vitamin D in order to illustrate some of the complexities involved. In addition, vitamin D status and recommended levels for vitamin D sufficiency will be discussed.

2.2.4.1 Photobiology, Photochemistry and Metabolism

It is not only chlorophyll in plants which photosynthesize sunlight: there are chemicals in human skin which photosynthesize light to produce vitamin D₃ and other reactants which play an important role in the immune system (Holick 2004a; Juzeniene et al. 2011; MacLaughlin et al. 1982; Norman and Bouillon 2010; Webb and Holick 1988).

Figure 2.7 illustrates some of the reactants and photoproducts which are involved in vitamin D₃ photosynthesis. The important point to note in this context is that the photosynthesis of vitamin D₃ by UV does not occur in isolation. There are a number of processes which are in quasi-equilibrium and the amount of previtamin D₃, and hence vitamin D₃, which is synthesized is limited; the equilibrium mixture contains, at most, 12–15% of previtamin D₃. The conversion of previtamin D₃ to vitamin D₃ is a slow, heat driven process (skin temperature over a period of hours or days) which can take place in the absence of UV exposure. Once synthesized, vitamin D₃ is taken into the blood to bind with vitamin D binding proteins (DBPs) which carry it to the liver. However, excess vitamin D₃ remaining in the skin may be broken down by further UV exposure (Webb et al. 1989). This, combined with the limited amount of previtamin D₃ synthesized, regulates the amount of vitamin D₃, and prevents vitamin D₃ intoxication from excessive over-exposure to UV.

Vitamin D₃ synthesis can also be limited by the amount of the precursor provitamin D₃ (7-DHC, 7-dehydrocholesterol) which is available. Although in ready supply in the skin, in adults the amount decreases linearly with age, with 80 year olds having around half that of 20 year olds (MacLaughlin and Holick 1985; Webb and Holick 1988). It might be assumed that this should be attributed to a reduction in the skin thickness with age reducing the amount of 7-DHC in the skin. However, MacLaughlin and Holick found that this is not the case—most of the 7-DHC is in the epidermis, the weight of which varies little with age. In contrast, in the dermis, which does decrease with age, the concentration of 7-DHC per unit area remains little changed. So although the

concentration of 7-DHC in older skin is reduced, which no doubt contributes to the reduced capacity of the skin to photosynthesize vitamin D, the reason for the reduction of 7-DHC is not due directly to a reduction in skin thickness.

The initial form of vitamin D₃ which is carried to the liver is biologically inert; it requires further hydroxylations in the liver to its 25(OH)₂D₃ form (which can be stored in fatty, adipose tissue) and then the kidneys and other tissues to become the 1 α ,25(OH)₂D₃ form which can react and bind with vitamin D receptors (VDRs). Any measure of the vitamin D blood serum level is generally reported as the level of circulating 25(OH)D, without differentiating between the photosynthesized vitamin D₃ form or the less potent vitamin D₂ form which is only acquired through diet (section 2.2.3.5). What is considered a sufficient level of circulating vitamin D will be discussed in section 2.2.4.2.

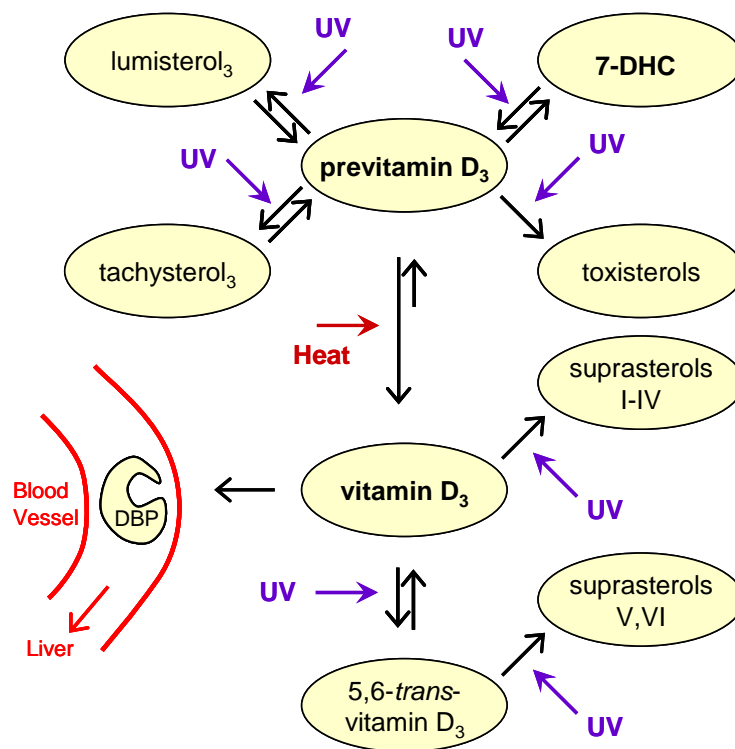


Figure 2.7 Photosynthesis of vitamin D in the skin

UV exposure to the skin drives several wavelength dependent conversions. Previtamin D₃ (precholecalciferol) is in quasi-photoequilibrium with its precursor provitamin D₃ (7-dehydrocholesterol, 7-DHC) and inert photoproducts lumisterol and tachysterol. Heat drives the conversion of previtamin D₃ to vitamin D₃ (cholecalciferol). Vitamin D₃ binds to vitamin D binding proteins (DBPs) in the blood and is carried to the liver. Excess vitamin D₃ is in quasi-photoequilibrium with 5,6-*trans*-vitamin D₃. These isomers are also converted to suprasterols (mainly suprasterols I and II). Toxisterols and other photoproducts (not shown) may also be formed. After Figure 3 of Holick (1994) and others (Chen et al. 2010; Green 1951; Jacobs 1995; Jacobs et al. 1981; Juzeniene et al. 2011; MacLaughlin et al. 1982; Norman and Bouillon 2010; Webb and Holick 1988).

For any photosynthetic reaction to take place, light of the appropriate wavelength must hit the precursor molecules. A schematic of the skin which also illustrates how far incident sunlight penetrates the various layers, is given in Figure 2.8. Note that the different wavelengths of sunlight, from the various ultraviolet (UV) constituents (UVA, UVB and UVC) through the visible to infrared (IR, heat), penetrate to different depths. So, if the processes within the different skin layers can be understood, and if the dependence of MS on the different components of sunlight can be elucidated, then it should be possible to isolate which processes are key to understanding the development of MS. For example, any UVA or UVB initiated processes within the skin must take place in just the upper layers of the skin. However, any processes which could be initiated by UVC are unlikely to occur since UVC does not penetrate the skin. Vitamin D photosynthesis takes place within the lower layers of the epidermis so any attenuation of incident UV by this and higher layers will restrict the amount of UV which can be used by vitamin D photosynthesis. The main restricting factor in human skin is thought to be the amount of melanin, which relates to skin type. Melanin acts like a sunscreen competing with 7-DHC for photons (Tsiaras and Weinstock 2011; Webb and Holick 1988), though the precise relationship between melanin and vitamin D synthesis needs further investigation since many results appear to be contradictory (Springbett et al. 2010). It is possible that the complex balance between the various

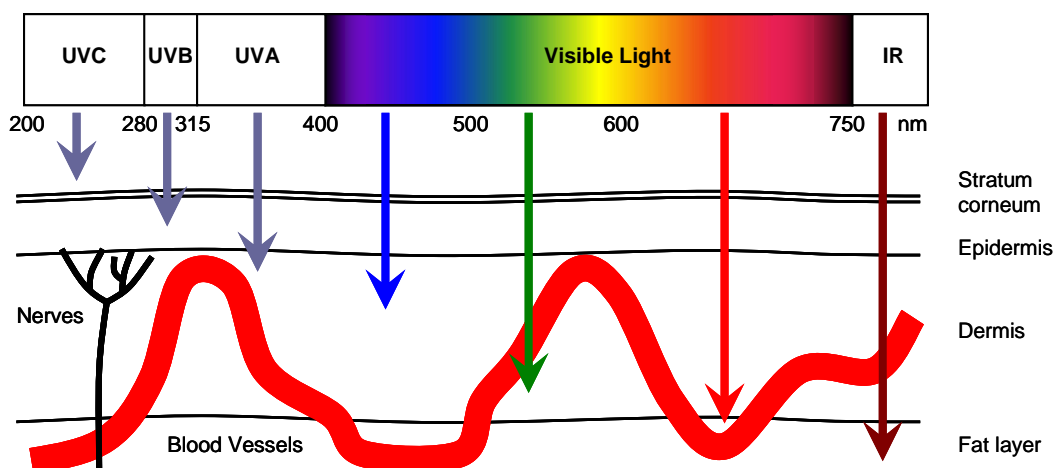


Figure 2.8 Penetration of incident sunlight to different depths of the skin
 After Figure 7 of Juzeniene (2011). The numbers give the wavelengths (nm) of the different constituents of visible and near-visible light. UVC does not penetrate the skin, UVB penetrates into the epidermis and UVA penetrates into the upper section of the dermis. Visible light penetrates to various depths through the dermis and infrared reaches the fat layer.

vitamin D related photoproducts has been overlooked as a serious confounding factor. Regardless of the finer details, it is established that darker skinned individuals photosynthesize vitamin D more slowly than those with a lighter skin for a given UV dose (Chen et al. 2007; Clemens et al. 1982), though the pigmentation does not appear to affect the amount of vitamin D synthesised, given enough time.

Attenuation of available UV is wavelength dependent, as are the various UV driven processes illustrated in Figure 2.7 and, hence, so is the equilibrium mixture of vitamin D related photoproducts. In addition, the amount of vitamin D taken into the blood depends on the initial 25(OH)D concentration (Edvardsen et al. 2007). These factors could combine to explain variations observed for UV thresholds below which vitamin D is not synthesised (Chen et al. 2007; Edvardsen et al. 2007; Engelsen et al. 2005; Slominski and Wortsman 2000; Webb et al. 1988).

Chapter 6 will discuss seasonal variations in UV levels, though, as mentioned above, for now it is sufficient to note that winter months see less sunlight than summer months. This also means that winter months see less UV radiation than summer months and so the amount of vitamin D which can be synthesised is reduced. At some latitudes there is no possibility of vitamin D synthesis for many months of the year—this is termed the 'vitamin D winter' (Edvardsen et al. 2007; Engelsen et al. 2005; Engelsen and Kylling 2005; Gillie 2010). At such times, vitamin D can only be obtained from dietary intake and stored reserves, which slowly release vitamin D back into the blood (Lagunova et al. 2009). It may be, therefore, that vitamin D is not released back into the blood at a sufficient rate during winter to sustain adequate vitamin D blood serum levels. It should also be noted that the concentration of vitamin D in stored reserves will be similar to that in the blood serum and that stored vitamin D is released in order to maintain the balance. So, although, individuals who are overweight or obese have greater quantities of vitamin D stored, the rate of release of vitamin D into the blood will only be slightly greater than that experienced in individuals who are either less overweight, or not overweight, and who have less vitamin D stored. That is, greater stored vitamin D will not prevent a decline in vitamin D status, it will only lessen the rate of decline. Vitamin D status is discussed further in the next section.

For additional details of the endocrinology of the skin, the reader is directed to Slominski and Wortsman (2000).

2.2.4.2 Status and Maintaining Sufficiency

Vitamin D has long been associated with various aspects of health (Juzeniene et al. 2011; Norman and Bouillon 2010). However, partly due to changes in lifestyles which have reduced the amount of time which people spend outdoors (Diffey 2006), the natural amount of sunlight which most fair skinned people are exposed to is generally insufficient for their needs (Glass et al. 2009; Glerup et al. 2000; Grant and Holick 2005; Holick 1994; Holick 2004a; Holick and Chen 2008; Ponsonby et al. 2010; Ponsonby et al. 2005; Rockell et al. 2006; Thompson 2007; Webb and Engelsen 2006; Webb et al. 2010; Zhang and Naughton 2010). This low exposure, combined with other factors such as diet discussed in section 2.2.3, leads to hypovitaminosis D, vitamin D insufficiency or vitamin D deficiency, and risk of disease (Dobnig 2011; Holick 2004b; Holick 2010), even in sunny countries, and in darker skinned populations (Oren et al. 2010; Unger et al. 2010). Indeed, even rickets, the classic vitamin D deficiency disease, is making a comeback (Holick 2006). Others conclude that people can obtain sufficient vitamin D through exposure to sunlight, even at latitudes such as the UK (Rhodes et al. 2010). However, this assumes that the recommended levels for a 'sufficient' vitamin D status are correct. Although some figures are well accepted, there is continual debate on what the correct values should be, and in general, official recommendations are well below those proposed or recommended by researchers in the field (Bolland et al. 2007a; Gilaberte et al. 2011; Livesey et al. 2007a; Rhodes et al. 2010; Scragg and Bartley 2007; Webb et al. 2010). Rovner and O'Brien (2008) reviewed the vitamin D status in US children. They started their summary by stating:

"No consistent definitions of vitamin D deficiency or vitamin D insufficiency were used, and several studies used multiple cut points for low vitamin D status based on 25-hydroxyvitamin D concentrations. Some studies reported 25-hydroxyvitamin D concentrations in nanomoles per liter, while others reported results in nanograms per milliliter. The values corresponding to vitamin D deficiency ranged from less than 5 ng/mL to less than 12 ng/mL, and for vitamin D insufficiency they ranged from less than 10 ng/mL to less than 32 ng/mL. In addition, there was considerable variability in the assay used to measure 25-hydroxyvitamin D concentrations ... there is often substantial variability among results from different assays." (Rovner and O'Brien 2008, p.515)

(Issues involved with assessing vitamin D status in population based studies are also discussed by Millen and Bodnar (Millen and Bodnar 2008).)

Reference	25 hydroxyvitamin D (25(OH)D) serum levels Thresholds between status levels in nmol/l [ng/ml]					
Brustad et al (2004b)	37.5 [15]		50 [20]			
Grant and Holick (2005)	50 [20]			80 [32]	250 [100]	325 ¹ [150]
van der Mei et al (2007b)	25 [10]		50 [20]	80 [32]		
Gozdzik et al (2008)	50 [20]			75 [30]		
Holick and Chen (2008); Holick (2009)	50 [20]		75 [30]			375 [150]
Rovner and O'Brien (2008)	[5–30]		[10–32]			
Rhodes et al (2010); Webb et al (2010)	12.5 [5] ²		50 [20]	80 [32]		
Vieth (2005)	25 [10]		40 [16]	75 [30]	160 [64]	220 [88]

Table 2.2 Blood serum levels of 25(OH)D for vitamin D status

Threshold levels of blood serum 25 hydroxyvitamin D (25(OH)D) reported for different vitamin D status levels in nmol/l and ng/ml (in brackets) in the cited references. See also references cited therein. Vitamin D status levels are given in italics. Notes: ¹ This is the figure reported. Comparisons with other figures in this table suggest that it should read 375. ² This is the figure used by the UK Department of Health as the level for vitamin D deficiency.

These inconsistencies are not confined to studies of children (Tsiaras and Weinstock 2011). Table 2.2 gives a few of the vitamin D levels reported elsewhere which also illustrates the additional issue of differences in terminology. For example, the terms 'sufficient' and 'optimum' are often used interchangeably, or together in the same report; similarly for the terms 'deficient' and 'insufficient'. For convenience, and without prejudice, the threshold limits for vitamin D status referred to in this and other sections will be taken as:

deficient < 50 nmol/l < insufficient < 75 nmol/l < sufficient < 80 nmol/l < optimum

or

deficient < 20 ng/ml < insufficient < 30 ng/ml < sufficient < 32 ng/ml < optimum

A further complication is that what might be a sufficient, or even optimum, vitamin D status during the summer months may not be adequate for sustaining a sufficient status through the winter without additional dietary intake (Bolland et al. 2007b; Burgaz et al. 2007; Gillie 2010; Gozdzik et al. 2008; Huotari and Herzig 2008; Lagunova et al. 2009; Livesey et al. 2007b; Rockell et al. 2006; Scharla 1998; Sharma et al. 2011; van der Mei et al. 2007b; Webb et al. 2010). The half-life of circulating vitamin D, including any released from fat stores, is around two months (Lin et al. 2011; Shroff et al. 2010; Tsiaras and Weinstock 2011; Vieth 2001; Webb et al. 2010). So, for example, even if an individual has an accepted optimum blood serum level of 90 nmol/l (36 ng/ml) at the start of winter, but then has no vitamin D synthesis from UV exposure, and no further dietary intake of vitamin D, then two months later they would have a blood serum level of only 45 nmol/l (18 ng/ml), which would be vitamin D deficient. This estimate compares well with the study of Webb et al (2010) who found that volunteers following their normal routines required late summer levels of 30.4 ng/ml (women) and 34.9 ng/ml (men) in order to have a February level of 20 ng/ml.

So what level of UV exposure should an individual aim for, or what should their dietary intake of vitamin D be, in order to ensure that their vitamin D status is maintained at either sufficient or optimum levels? Dietary requirements will be discussed below, followed by UV exposure levels. First, however, it is necessary to define a few more units of measurement.

Blood serum levels of circulating 25(OH)D are given in either nmol/l or ng/ml, as discussed above. Oral vitamin D doses are generally specified in International Units (IU, 100 IU = 2.5 µg). For exposure to UV radiation (UVR), reference is often made to a minimal erythemal dose (MED) where 1 MED is defined as the lowest dose of UVR which causes faint pinkness or reddening of the skin within 24 hours (Tsiaras and Weinstock 2011; Webb et al. 2011b). The length of time required to achieve 1 MED varies with geographical location, skin pigmentation, percent body fat, and age (Grant and Holick 2005). The actual dose received by an individual also depends on the amount of body surface exposed to the UVR. More recent publications report UVR exposure expressed in standard erythemal dose (SED) units, where 1 SED = 100 Jm⁻² of erythemal UVR (McKenzie et al. 2009; Webb et al. 2010; Webb et al. 2011b). The ratio of SED to MED gives a pigment protection factor which relates to skin type (Bogh et al.

2011). In addition, there is the biologically effective dose (BED) (Edvardsen et al. 2007) which estimates a true vitamin D dose by considering true incident UV, weighted by the vitamin D action spectrum (see also section 6.2.2) and the standard vitamin D dose (SDD) corresponding to a UV dose equivalent to an oral dose of 1000 IU.

As with blood serum levels discussed above, it is now accepted by researchers in the field that official recommendations for dietary intake of vitamin D are too low (if UV exposure is limited), possibly by an order of magnitude (Glerup et al. 2000; Heaney et al. 2009; Holick 2009). It is now believed that a daily intake of 1500–2000 IU is required for adults to *maintain* a sufficient/optimum level of vitamin D (Holick 2009), with even higher doses required in the absence of UV exposure (Gillie 2010). In addition, upper limits for safe vitamin D doses appear to have been over cautious, based on reports of a few extreme cases of overdoses, and the persistence of misdiagnoses (Holick 2009; Kimball et al. 2007; Vieth 2006). There is emphasis, above, on the word 'maintain'—if the individual is vitamin D deficient then these doses will not elevate the blood serum level of vitamin D to optimum or even sufficient level. To treat vitamin D deficiency it is first necessary to boost the vitamin D intake to bring it up to sufficiency before then starting a maintenance regime (Adams and Hewison 2010; Holick 2009). An extensive list of strategies to treat vitamin D deficiency are given in Table 3 of Holick (2009). In addition, he reports that:

"To treat vitamin D deficiency, I typically give my patients 50,000 IU of vitamin D₂ once a week for 8 weeks to fill the empty vitamin D tank followed by 50,000 IU of vitamin D₂ every 2 weeks. I have followed patients up to 5 years on this regimen and most patients have blood levels between 30 and 50 ng/ml within 2–3 months after initiating treatment. For patients who have a blood level of 25(OH)D of >30 ng/ml in order to maintain this level, I give my patients 50,000 IU of vitamin D₂ once every 2 weeks." (Holick 2009, p.13)

Calvo et al (2005) reviewed the dietary intake of vitamin D based on reports from many countries including Canada, Japan, Norway, Spain, the United Kingdom, and the United States (US). Mean daily intakes for adults ranged from 132 IU for Spanish women to 325 IU for Caucasian US men (women 292 IU). Even with food fortification and the use of supplements, US adults only consume a fifth of the lower limit of 1500 IU recommended by Holick and others. Further, it is now estimated that the body's requirement for vitamin D is 3–5000 IU per day, which is several times higher still (Heaney et al. 2009; Holick 2009).

In the absence of any dietary intake of vitamin D, individuals must rely on UV exposure to synthesise vitamin D, as described in section 2.2.4.1. However, due to the regulatory mechanisms in place, a given dose of UV exposure will result in different increases in circulating vitamin D in different individuals. The main regulatory factor is the current blood serum vitamin D level of the individual—UV exposure and subsequent vitamin D synthesis will only 'top up' the current level. Despite this, it is possible to estimate requirements for various current status levels, using controlled exposure studies. Holick has estimated that UV exposure of 1 MED over the whole of the body's surface area is equivalent to an oral dose of 10000–25000 IU vitamin D (Holick 2004a; Holick 2009; Webb et al. 2011a). However, this vitamin D dose is unlikely to be anything close to any 'real life' exposure since few are likely to expose their whole body to 1 MED of UV radiation intentionally, and the regulatory processes would limit the amount of vitamin D synthesised.

Webb et al (2011a) explain how results from controlled experiments, such as from their earlier paper (Rhodes et al. 2010), can be related to real life situations by modelling the real life exposure which an individual wearing knee-length shorts and T-shirt might experience when standing and moving around. (The authors consider a number of solar elevation scenarios; for simplicity, only one is described here.) A UV exposure of 1.1 SED (0.44 MED for skin type II (McKenzie et al. 2009)) was received from 6.5 mins of 35% full body exposure in a solar cabinet. In Manchester, UK (53.5° N), with a noon solar elevation of 60°, the same exposure could be received by a person moving around for 33 mins, or lying flat for 26 mins (13 mins each side). It may be, however, that this estimate of 33 mins is an overestimate. Although the authors consider different (vertical) angles to the sun, and note that when the surface considered is facing more than 90° away from the sun it only receives scattered radiation, they do not appear to consider the simultaneous exposure of the *back* of the surface which would then be facing the sun (or vice versa when the surface faces the sun). Although the area of a person's face would only present to the front, someone wearing shorts and T-shirt will present much the same arm plus leg area to both front and back, unless the arms were held across (or behind) the torso. (There would also be small differences due to changes in angles when bending arms or legs when walking rather than standing upright.) It is possible, therefore, that someone walking around may receive a given UV exposure dose quicker than someone lying sunbathing

(assuming the same clothing). Regardless of the exact time required for the 1.1 SED, Webb et al estimate that this dose would be equivalent to an oral vitamin D dose of 2000 ± 600 IU, which is the upper limit of the 1500–2000 IU per day range given above for the vitamin D intake required to maintain a sufficient or optimum blood serum level. In their earlier paper (Rhodes et al. 2010), the authors also note that this 1.1 SED dose, which was administered three times a week for six weeks, was able to raise the participants' blood serum level up to a 'sufficient' (20 ng/ml), but not 'optimum' (32 ng/ml) level, with little increase in the final week suggesting that a plateau had almost been reached. However, the 20 ng/ml level referred to as sufficient would only be 'insufficient' using the limits adopted here. The same UV exposure daily rather than three times a week would be required to give the equivalent of a daily dose of vitamin D of 2000 ± 600 IU.

Note that incident UV radiation is highly dependent on solar elevation, as will be discussed in Chapter 6, and very little UVB radiation reaches ground level at low solar elevations. This means that at some latitudes, even though there may be several hours of daylight in winter, there is little or no useful UV radiation for synthesis of vitamin D. (See also Appendix F for the effect of solar elevation.) In addition, these figures assume a clear sky; the presence of clouds will reduce the dose received considerably, as also illustrated in Chapter 6.

In summary, to maintain an optimum level of circulating vitamin D, a regular dose equivalent to a daily dose of 1500–2000 IU is required. This can either be obtained directly from diet or supplements, or from UV radiation exposure. The duration of UV exposure required varies depending on solar elevation, but at noon in mid summer in Northern England 26–33 mins would be sufficient to provide 2000 IU. In the context of the study here, and based on data for all of the solar elevations (30–60°N) from Webb et al (2011a), New Zealand residents would require mid summer exposure around noon of 21–27 mins in the north and 24–30 mins in the south.

2.2.5 Timing of UV Exposure in MS

Interestingly, the aetiology of MS appears to be dependent on exposure to sunlight at different times of life, not just pre-birth. There are three main periods of concern: prenatal, early childhood and prior to onset. These will now be discussed in more detail.

2.2.5.1 Prenatal

The most likely process which could have a role in MS aetiology during the prenatal period is the formation of the central nervous system (CNS) during the embryonic and foetal stages of development. Given that MS manifests itself as a demyelination of the nerve cells it seems possible that something is acting to interrupt or damage this process. Some now believe that MS is a neurocristopathy (Behan and Chaudhuri 2010; Behan 2010). That is, that something affects the neural crest cells which form the oligodendrocyte cells involved in myelin sheath formation (Bolande 1974; Knecht and Bronner-Fraser 2002). (See Figure 2.9 for an illustration of the neural crest formation.) It is proposed that some (unspecified) weakness or predisposition to damage is introduced which then manifests itself at a much later stage in the MS pathology. At what stage this might occur, and whether this weakness would affect just the oligodendrocytes which form the myelin sheaths or also the other glia cells created from the neural crest, is not yet clear, though it is clear that oligodendrocyte cells are

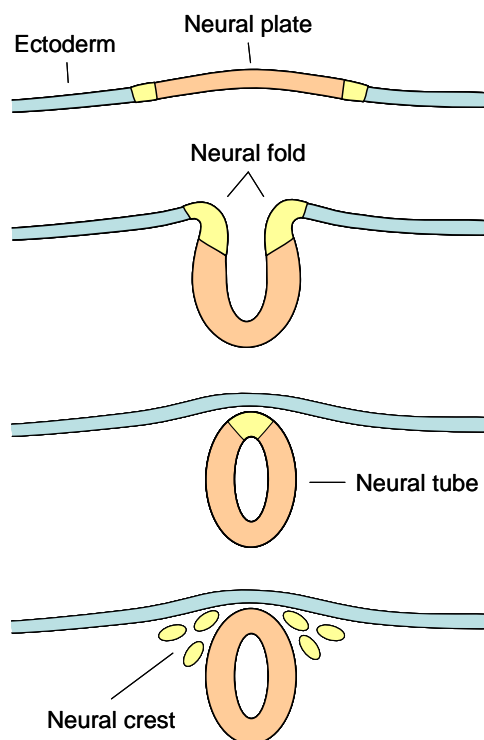


Figure 2.9 Schematic of neural crest formation

The neural tube forms from folds in the ectoderm during the first month of embryonic development. Later (not shown here), the neural crest cells differentiate to form other sensory neurons and glia cells such as the oligodendrocytes and schwann cells of the central and peripheral nervous systems (CNS and PNS) which are involved in myelin sheath formation. After Knecht and Bronner-Fraser (2002).

also involved in remyelination and, possibly, the failure of the remyelination process following demyelination (Imitola et al. 2003; Kotter et al. 2011). Whilst not disagreeing with the suggestion that MS is a neurocristopathy, Krone and Grange (2011) point out that this must be considered in conjunction with other immuno-regulatory factors and not in isolation.

As to what external factor might influence the embryonic and foetal stages of CNS development, the main candidate is still a combination of vitamin D and UV exposure. Low levels of maternal vitamin D are well established as links in various diseases (Juzeniene et al. 2011; Ponsonby et al. 2010), though the critical stage of pregnancy does vary by disease. For MS, the critical period appears to be the first and early second trimesters. Staples et al (2010) found a strong month of birth effect when looking back at data collected across Australia in 1981. Specifically, the authors found a strong inverse association between ambient UV levels during the first trimester of pregnancy and MS, which persisted after adjusting for various factors, including gender and region of birth. In addition, the month of birth effect did not persist after adjusting for ambient UV levels during the first trimester, suggesting that UV levels during the first trimester were sufficient to explain the observed month of birth effect. Although the association was strong for the first four or five months of pregnancy, the strongest association was seen to be with UV levels during the second and third months, that is, after the neural crest has formed and corresponding to the period of greatest production of neurons and greatest development of the CNS, though before any myelination of neurons occurs (Hepper 2006). However, any lag between low incident UV and low vitamin D status would need to be allowed for when considering which development processes might be affected. This lag could be up to several months in winter and spring.

2.2.5.2 Childhood and Adolescence

Early childhood is the period from birth while the body is growing and developing fast, however, growth continues throughout childhood and adolescence. It is likely that factors which affect the formation of the myelin sheaths will still be factors after birth, and other processes will still be vulnerable throughout this period of growth. However, there are now more environmental factors to consider as the child is no longer in the mother's womb and its own immune system responses will not start to develop for

several months. The child will gain some antibodies from the mother's immune system through breast milk, however, that may be of mixed blessing. Although it provides the child with protection against many illnesses, it could also pass problems with the immune system down from mother to child. In fact, the immune system has long been accepted as key to understanding MS, so an inheritance by the child of elements of the mother's immune system could, in part, confound any genetic familial factors. MS is no longer considered by many to be an auto-immune disorder as the immune system does not appear to be reacting to any specific antigen (Rodriguez 2009). However, accepting that the immune system does play a major role, the main question now is whether the immune system is somehow causing the damage to the myelin sheaths or whether something else is causing the damage and the immune system cannot repair the damage. (Or a combination of both.) Given that repair is possible (Stangel et al. 2011), it would appear that there is some balance going on (Krone et al. 2011).

The effects of UV exposure during childhood and adolescence on the prevalence of MS are best illustrated by studies which look at the effects of migration. Migration is a confounding factor for other effects such as variations in UV exposure due to latitude of residence. If these factors are allowed for, then the effects of migration at different ages can indicate the effects of UV exposure at different ages. Some of the earliest studies to look at migration considered MS within immigrants to Israel, most of whom had come from areas of Europe with higher levels of prevalence of MS than that seen in the non-immigrant Israeli population (Alter et al. 1978; Alter et al. 1966). However, rather than having the same risk of MS as non-immigrants, immigrants migrating after the age of 15 were found to keep the risk associated with their country of origin. Only immigrants migrating before the age of 15 acquired the risk associated with Israel. Similar results have since been observed in other immigrant populations (Kurtzke 2000) including those to Australia (McLeod et al. 2011), Canada (Kennedy et al. 2006), Norway (Smestad et al. 2008), the United Kingdom (Elian et al. 1990), within the United States (Visscher et al. 1977; Wallin et al. 2009) and return migration to the West Indies from France (Cabre et al. 2005), though the age threshold for migration effects was presumed to be 15 within some studies, and not investigated explicitly within others. This body of evidence points to a strong protective effect of UV exposure during childhood and adolescence.

2.2.5.3 Prior to Onset of MS

If an individual inherits a susceptibility to MS, and damage is caused prior to birth and during childhood and adolescence, then it is likely that MS as a disease is already established before onset is observed. Onset then results either due to some trigger, which causes symptoms to occur above the clinical threshold (Ramagopalan et al. 2010) , or else some protective mechanism is reduced or removed, which had been preventing symptoms from occurring above the clinical threshold. The immune system is in a state of balance. If that balance tips too far and nerve damage cannot be repaired then symptoms of MS will follow. How the disease progresses will then depend on how severe the damage is and how much of the damage can be repaired. It is thus tempting to consider that there may not be any individual trigger and onset of MS is due solely to the reduction or removal of some protective mechanism which allows the immune system balance to tip. But what is that protective mechanism?

A serious candidate for a protective role in the pathology of MS is vitamin D. So if vitamin D levels have an effect on this immune system balance, as is believed, then, following on from the various discussions in section 2.2.3 and 2.2.4, it is likely that this will translate to an observed effect on MS prevalence rates due to the level of UV exposure. However, the reverse is not necessarily true. Lucas et al (2011b) report that both sun exposure and vitamin D levels are independent risk factors for first CNS demyelinating events, a precursor to MS, and so both factors may have independent roles in the pathology of MS. The effect of UV exposure independent of vitamin D has also been highlighted or studied by other authors (Artuković et al. 2010; Bains 2010; Beretich and Beretich 2009; Gong et al. 2008; McMichael and Hall 1997; Ponsonby et al. 2002; Simpson et al. 2011; Sloka et al. 2011; van der Mei et al. 2003; Warren et al. 2011), though many other studies have concentrated on the connection between vitamin D and MS without regard to an effect of UV exposure and MS independent of vitamin D (Ascherio and Munger 2007b; Kampman and Brustad 2008; Kampman and Steffensen 2010; Mehta et al. 2011; Pierrot-Deseilligny and Souberbielle 2010; Smolders et al. 2008; Smolders et al. 2011).

2.2.6 UV Exposure—Benefits and Risks

Despite the fact that it is visible light and infrared radiation which penetrate the skin furthest, it is the UV elements of light which are believed to play the most critical role

in the aetiology and pathogenesis of MS. The most well-known product of UV photosynthesis in the skin is vitamin D, as described above, and it has long been associated with various aspects of health (Juzeniene et al. 2011; Norman and Bouillon 2010).

The extent of vitamin D deficiency, described in 2.2.4.2, does not bode well for MS prevalence since vitamin D also plays a major role in MS, though this role is not straightforward: vitamin D appears to assist in suppressing certain auto-immune mechanisms while at the same time mediating other anti-inflammatory processes (Holick 2004a; Ponsonby et al. 2002; Ponsonby et al. 2005; Smolders et al. 2008; Smolders et al. 2011).

Whereas exposure to UV is good for vitamin D synthesis, extended exposure to UV, leading to sunburn, is associated with skin cancer. This has prompted campaigns in countries such as the UK and Australia to persuade people to limit the potential for getting sunburnt (AUDHA 2010; CCA 2012; CRUK 2009, 2012; Gillie 2010; Hiom 2006; van der Mei et al. 2007b). In countries with high UV exposure, such as Australia, the result is a tendency to stay out of the sun or to cover up, when reminded to do so (Dobbinson et al. 2008), which may have helped to reduce or stabilise the incidence of skin cancer, at least in younger age groups (AUAH 2011; Sinclair and Foley 2009). Incidental UV exposure should be sufficient to keep vitamin D levels at a sufficient level (Marks et al. 1995), though not always so in winter (Czarnecki et al. 2009). The situation in the United Kingdom (UK) appears to be somewhat different—with a latitude of 50–61°N UV levels in the UK are not always high enough to enable synthesis of vitamin D, especially in winter (Engelsen et al. 2005; Thompson 2007; Webb et al. 2010). There may also be some confusion regarding the advice people should be following. Gillie (2010, Table 1) points out that the Cancer Research UK recommendation to "stay in shade between 11 am and 3 pm" changed in 2006 to "spend time in the shade between 11 am and 3 pm". Given that the window of opportunity for vitamin D synthesis is limited in the UK, anyone following the pre-2006 advice is likely to restrict their ability to generate vitamin D through sunlight (Rhodes et al. 2010). However, there is still a tendency for many British adults to seek out the sun in order to obtain a tan (CRUK 2007, 2008). As Hiom (2006) states: "the sporadic availability of prolonged sunny spells makes our sunshine a rare commodity that many feel must be 'made the most of', whether within the UK or abroad. It would also

appear that this is a European attitude rather than just a British one—holiday makers throughout Europe are prone to the same behaviour (Argyriadou et al. 2005; Ezzedine et al. 2007; Køster et al. 2011).

So while there is a general tendency to stay out of the sun, possibly contributing to vitamin D insufficiency or deficiency, when time is intentionally spent in the sun, it is more often prolonged or under extreme conditions, which can lead to sunburn and skin cancer. In addition, the high sun protection factor (SPF) sunscreens which are used by many intentional sunbathers are often applied too thinly so that users do not receive the UV protection which they expect (Autier et al. 2007; Ezzedine et al. 2007; Norval and Wulf 2009). Meanwhile, individuals expecting only casual exposure to sun, may use lower SPF sunscreens, which are also more likely to be used correctly. An SPF 8 sunscreen will reduce the UVB radiation incident on the skin by up to 95% (Holick 2004a; Holick et al. 2007), so an individual using an SPF 8 sunscreen may reduce the potential of vitamin D synthesis by a factor of 20. In addition, although the use of a sunscreen may suggest that the user will spend longer in the sun, it is possible that only durations of intentional sun exposure are extended, with casual exposure being unaffected (Autier et al. 2007).

It can be seen that there is a balance between too much and too little exposure to sunlight which varies by location. The balance is a delicate one which has led to differences of opinion on the health care advice which should be given out, and what various government policies should be (Diffey 1998, 2006; Gillie 2006, 2010; Hiom 2006; Kimlin et al. 2007; Macdonald et al. 2010; McKenzie et al. 2009; Moan et al. 2008; Rhodes et al. 2010). It is also apparent that recommended dietary intake of vitamin D needs to be clarified and realistic estimates of intake taken into consideration when discussing UV exposure.

2.3 Summary of MS Epidemiology

MS is a complicated disorder, or it may even be several disorders. The underlying susceptibility to MS appears to be multigenic through a complex interplay of different genetic factors. Whether a susceptible person then develops MS and goes on to show clinical symptoms of MS depends on a delicate balance between many genetic and environmental factors and processes, acting from pre-birth and throughout the person's life. Important genetic factors include, though are not restricted to, HLA

alleles and haplotypes associated with the immune system. Important environmental factors include diet, smoking and viral infection. However, the main environmental factor driving many of the processes appears to be exposure to sunlight, and UV radiation in particular, which is important in determining vitamin D status in the absence of adequate dietary intake. Although the detailed aetiology and pathogenesises of MS are not known, any better understanding of how UV exposure, through the surrogate latitude, affects prevalence rates, will help to clarify what the pathogenesises are and what might be done to improve the paliative treatment of MS, and ultimately, assist in finding a cure. In addition, links between sunlight, UV and early development will help to establish the MS aetiology with the aim of finding some means of preventing MS.

This study aims to add to what is already known about the effects of UV on MS prevalence rates by further analysing data collected in New Zealand (Taylor et al. 2010b). Details of the NZ study are given in the next chapter which discusses various elements of the geography of this study, along with other current knowledge of the geography of MS.

Chapter 3

The Role of Geography in Understanding MS

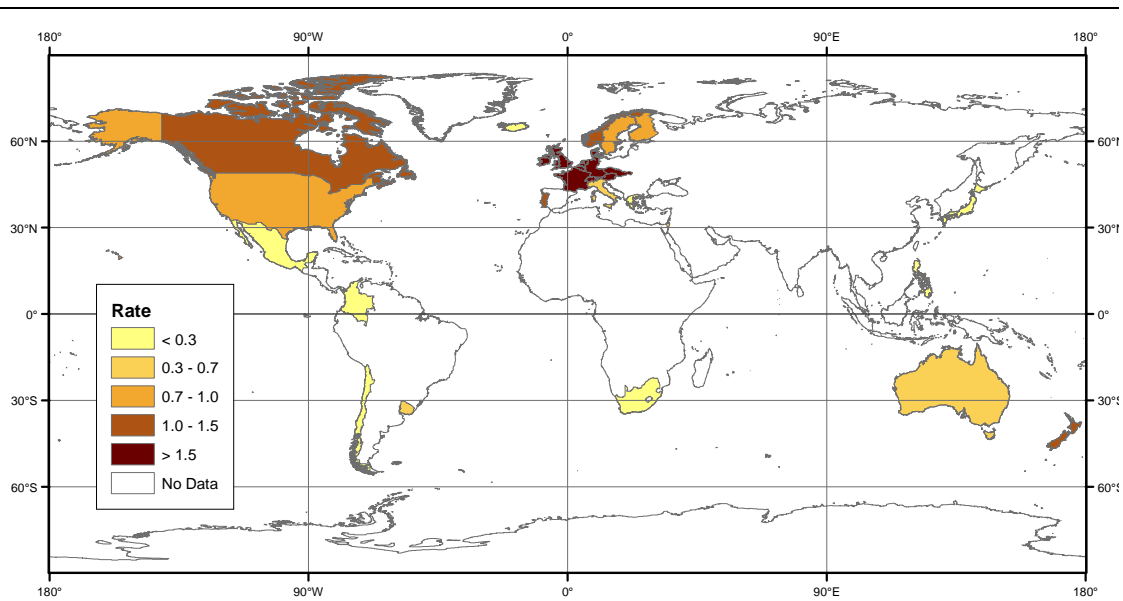
3.1 Introduction

This chapter describes the geography of MS, including the New Zealand MS Prevalence Study which has provided the data for this study. The data collected by the study is summarised, together with initial results of relevance here. Limitations of the data collected, and consequences of those limitations, are discussed. This chapter also describes the Modifiable Areal Unit Problem and consequences for studies on MS.

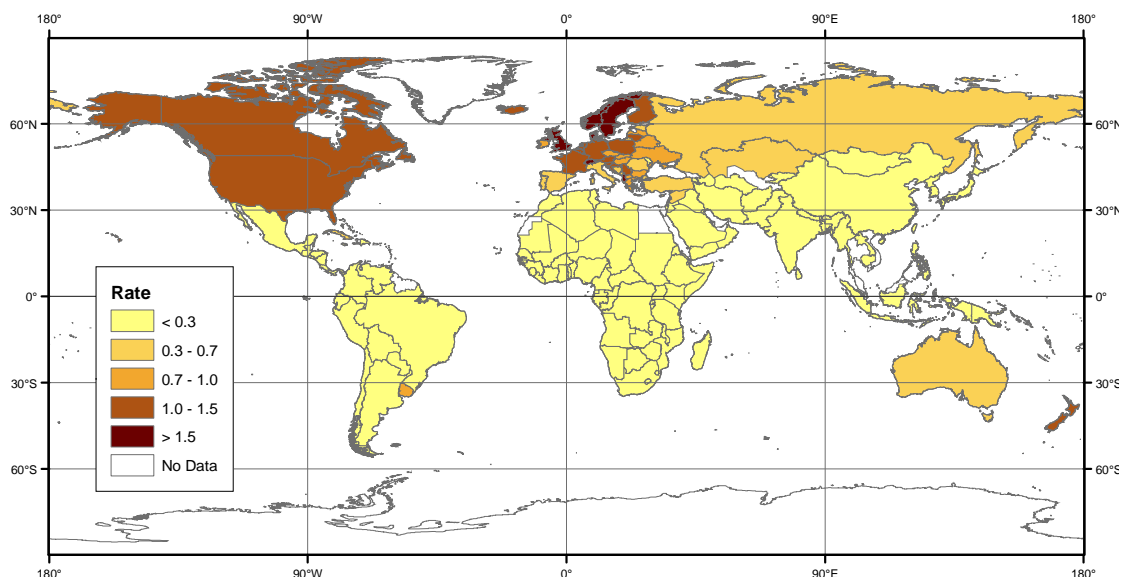
3.2 The Geography of MS

As mentioned in Chapter 2, two of the main factors which are associated with variations in the rates of prevalence of MS are ethnicity and latitude of residence, both of which are strong geographical factors, as discussed by Alter and Harshe (1975) who also noted the confounding effect of migration on prevalence rates. Most of the studies from the 1960's to 1980's which have looked at the prevalence of MS have concentrated on individual areas, towns and cities within populations of European ancestry since these were the populations seen to have the highest rates. In addition, these populations tended to be in countries with well developed health services where the collection of data could be carried out and the data obtained could be considered reliable. However, studies were often localised and there was no guarantee that case ascertainment and diagnostic criteria were consistent with other studies, a problem which still persists in many more recent studies (Kurland et al. 1965; Polman et al. 2011; Rosati 1994). For data from less localised studies, there was a further issue: prevalence of MS was often only given as a ratio based on the numbers of neurological patients admitted to hospital (Kurland et al. 1965). Only for smaller countries, such as Israel, were national studies of MS prevalence rates possible (Alter et al. 1978; Alter et al. 1966; Antonovsky et al. 1965). An alternative, possibly more reliable, source of data for an indication of true prevalence in situations where prevalence data is lacking or unreliable is that of recorded deaths attributed to MS. However, death rates will be closer to estimates of incidence rather than prevalence, and as such do not indicate the burden of MS on the area concerned. Also, as will be discussed, even these rates can be confounded by other factors.

Figure 3.1(a) illustrates the age-adjusted death rates for the mid 1950s, based on data taken from Kurland et al (1965) and dating from 1951 to 1958. The rates are age-adjusted to the 1950 US population and are per 100,000. The lack of data for most countries of the world is apparent, as is the strong correlation between data which is available and countries either within Europe or those colonised by early European settlers. Figure 3.1(b) shows the situation in 2008, based on data downloaded from the



(a)



(b)

Figure 3.1 Age-adjusted death rates for MS by country

(a) Rates by country per 100,000 population for the mid 1950s, age-adjusted by the 1950 US population. Data from Kurland et al (1965). **(b)** Summary estimated rates by country per 100,000 for 2008, age-adjusted by the WHO global standard population. (WHO 2011).

World Health Organization (WHO) data repository (WHO 2011). There are differences, as might be expected 50 years on, and not just through the better availability of data, which now covers most countries of the world, and the change of standard for the age standardisation. (The choice of standard in age standardisation is discussed in Chapter 4.) This data represents the rates of deaths attributed to MS and these rates could be affected by a number of factors. In particular, better health care could mean that more people may die of other causes before death attributed to MS; or they may survive other causes and so finally have death attributed to MS. Similarly, better awareness, understanding and diagnostic techniques might attribute more deaths to MS, or attribute more deaths to other neurological causes rather than MS. That the overall scale of death rates in the 1950s are so similar to those of 2008 suggests that these confounding factors broadly balance out. The global picture remains much the same—the highest rates are in parts of Europe, Canada, USA and New Zealand, with the countries between 30° north and 30° south having the lowest rates. Australia spans the 30°S meridian and is seen to have an intermediate rate.

3.2.1 Genetics or Latitude?

So what are the factors which determine the global prevalence of MS? Figure 3.1 suggests that there is a correlation between MS and distance from the equator. Yet the countries with the highest rates also have white European ancestry, and that suggests that there may be genetic factors involved. Establishing a global view on the prevalence of MS has not been easy. This section describes some of the problems encountered and continues by outlining many of the findings.

Although death rate data does not translate directly to prevalence rates, it does give a global picture of the geography of MS based on more consistent criteria, where prevalence rates are often lacking. Despite this, more localised prevalence studies are invaluable for teasing out an understanding of the different factors which are thought to be involved. Leibowitz et al (1967) combined data from two other sources and looked at correlations between MS prevalence rates and various geographical and socio-economic factors. It should be noted that one data source was of MS prevalence rates based on country of birth of immigrants to Israel, so migration will also be a confounding factor (Alter et al. 1966). Bearing this in mind, it is interesting that the strongest correlation found was a negative correlation with mean hours of sunshine,

which had a stronger correlation than that of latitude. However, the study also found strong correlations with various economic variables, including the consumption of steel, literacy and the numbers of physicians and dentists. As the authors pointed out, many of these indicators are indicative of the technological development of an area and so could reflect greater urbanisation, higher pollution, greater levels of sanitation, and better health care, amongst other factors. They went on to state:

"Of course, the mere finding of a correlation between the prevalence of MS and another variable is no proof that they have any causal connexion. In fact, for most of the variables studied, correlations were expected *a priori* because the variables were known to vary with latitude. It is not difficult to select factors which are related to known or suspected differences between groups with high and low prevalence rates of MS. Therefore, it would be unwarranted to draw broad conclusions from the data presented." (Leibowitz et al. 1967, p.884)

A major complication with MS is that all too many factors can be correlated with its prevalence. Therefore, in order to understand any particular factor, it is necessary to try to isolate other confounding factors, as mentioned in section 2.2.2. An attempt at modelling the variation in prevalence with latitude by Alter et al (1973) did confirm a strong correlation but was unable to provide a simple functional form which fitted the data at all latitudes. Looking at a subset of the data, however, was more successful: an exponential fit was seen to model the North America data much better than any other fit, either to just the North America data, or to the data as a whole. In this analysis, for prevalence of MS, the North American population can be considered as quite homogeneous both genetically (predominantly of a similar white, European ancestry) and culturally (similar standards of living and diet). As such, the main factor which is likely to affect the prevalence rates within North America is latitude and most other factors will be uniform within this sample. Ironically, within Europe itself, the situation is not as clear as it was for Alter et al for the North America data, as different sub-populations, such as those in the north of Scandinavia and in Sardinia, confound many latitude effects (Alstadhaug et al. 2005; Dahl et al. 2004; Krökki et al. 2011; Kurtzke 1977; Pugliatti et al. 2001b; Sundström et al. 2001).

As discussed in section 2.2.3, Harbo et al (2007) confirmed a prevalence of MS in the Norwegian Sami which was even lower than that found in the non-Sami Norwegian populations. They also found differences in the HLA frequency distributions between the two populations. Although complicated by possible dietary differences, this

suggests that the Sami have a genetic resilience to MS which dominates any effect due to latitude. However, in contrast to the Sami, the Sardinian genetic make up appears to have rendered the Sardinians susceptible to MS, rather than protecting them against it, with a complicated association with TNF- α alleles being key and dominating over any latitude effects (Marrosu et al. 2001; Sotgiu et al. 2004), as also discussed in section 2.2.3. In addition to the overall high prevalence of MS in Sardinia, there are also local variations in the prevalence (Pugliatti et al. 2009; Pugliatti et al. 2002). In general, these can be attributed to either distinct sub-populations where isolation, due to terrain, or intra- and interbreeding in the last century, have strengthened or weakened the Sardinians overall genetic susceptibility to MS, or to areas where other biological factors, such as malaria or virus outbreaks, especially after World War II, may have had an influence.

The prevalence rates of MS in Sardinia and the Sami are likely to be due to founder effects. These are often seen in small, or typically island, populations which exhibit limited immigration or interbreeding with other populations, and where the gene pool remains relatively isolated. Founder effects have also been observed in the Gypsy population of southern Spain (Fernández et al. 2008), Čabar, Croatia (Perković et al. 2010) and in Newfoundland and Labrador, Canada (Sloka et al. 2005a).

An extensive review of the global prevalence of MS is given by Rosati (2001) which collates data by country or country group with a brief discussion of temporal trends and any other factors worthy of note for the given data. It was also noted that, at the time of publication there were over 400 previous publications on the prevalence of MS throughout the world, yet the problem of defining the geographic behaviour of MS still remained. The latitude situation and potential founder effects in Europe were summed up in the abstract with the statement:

"The updated distribution of MS in Europe, showing many exceptions to the previously described north-south gradient, requires more explanation than simply a prevalence-latitude relationship. Prevalence data imply that racial and ethnic differences are important in influencing the worldwide distribution of MS and that its geography must be interpreted in terms of the probable discontinuous distribution of genetic susceptibility alleles, which can however be modified by environment. Because the environmental and genetic determinants of geographic gradients are by no means mutually exclusive, the race versus place controversy is, to some extent, a useless and sterile debate." (Rosati 2001, p.117)

The follow on review by Pugliatti et al (2002) groups data by continent and provides greater discussion of the various findings. The review starts, however, by highlighting some of the ever present difficulties when comparing prevalence data. These include: different age structures and population bases; mixed ethnicity; varying diagnostic criteria, case ascertainment, medical care and other health care resources. Time is another confounding factor—if prevalence rates vary with other factors which themselves vary with time, then it is difficult to compare prevalence rates taken at different times. Significant factors which affect MS prevalence rates and which vary with time include diagnostic criteria and resources, such as access to MRI scanners, and possible environment factors such as diet, pollution and UV exposure. No attempt is made here to reproduce the findings of either of these reviews.

Despite the problems when comparing different MS prevalence data, there is still a need for prevalence studies. Where significant factors exist, stratification of the data by these factors should allow the results to be compared with other data.

Alternatively, studies can be undertaken within populations where certain factors are absent. However, with an increase in global migration and the continued mixing of various gene pools, this isolation approach is fast becoming untenable. Additionally, any attempts at treatment or prevention of MS within a population, such as with dietary supplements, will introduce other confounding factors which will need to be taken into consideration in further studies within that population.

Since the reviews of Rosati (2001) and Pugliatti et al (2002), there have been well over 100 further studies published on MS prevalence rates in various locations around the world. Some of these studies are very localised, such as the study in Linguaglossa, Sicily (p. 5422) (Nicoletti et al. 2005); others are national such as that in France (p. 52m) (Fromont et al. 2010). Reasons for this explosion in data are many, but include:

- The timescales involved in the pathology of MS mean that longitudinal studies take time—studies initiated decades ago, or studies based on decades of data, are only now coming to fruition (Granieri et al. 2007; Sarasoja et al. 2004);
- There has been a wider acceptance that MS is not just a European disease, or one only affecting populations of higher latitudes. Studies based in countries with non-European populations, or at lower latitudes, are now being reported (Abad et al. 2010; Alter et al. 2006; Díaz et al. 2011; Houzen et al. 2003).

It must be kept in mind, however, that countries with low rates of MS may have less expertise in the treatment and diagnosis of MS, and countries which do not have infrastructure available to enable a comprehensive survey may be reliant on incomplete data. Both of these scenarios have implications on the availability, reliability and reporting of data. The recent study by Hasham et al (2010) on rates of MS in Egypt is more than welcome in that it attempts to fill in a gap in the global map. However, the study is only able to report findings based on the numbers of other neurological patients at a selection of referral centres, not exact prevalence, since the infrastructure does not exist in Egypt to enable a more comprehensive survey.

3.2.2 MS Prevalence Studies Going Into the 21st Century

Many more studies are now aimed at genetic factors rather than pure latitude effects, studies made much more viable with advances in genetic analysis techniques, as described in section 2.2.3.2. Table 3.1 lists recent studies of MS prevalence rates in Italy, the country with most new studies. Note that most of these are in Sardinia and Sicily. Both of these Italian islands have MS prevalence rates higher than might be expected from other Italian studies and relatively isolated genetic populations which suggest founder effects. They are also small enough that any variations in rates observed within each island are unlikely to be due to latitude effects and so must be due to some other environmental or genetic factors.

Recent studies reported for the rest of Europe are given in Table 3.2. As with much of the earlier data, many of the studies have been in Scandinavia. There is a study by Harbo et al (2007) which looks at the Sami population of Northern Norway. The Sami were thought to be somehow immune to MS, yet a few cases of MS are now being reported, as described in section 2.2.2. Also of note is the study by Fernández et al (2008) which looks at the Gypsy population in southern Spain, an example of the founder effect, described above. The Gypsies, or Romany people, are descended from the Gypsies of Eastern Europe, who in turn migrated from northern India. In addition to having this rather unique genetic background, cultural isolation has tended to limit any interbreeding with other peoples.

Country	Location (Study Group)	Study Years	Prevalence /10 ⁵	95% CI /10 ⁵	Reference
Italy	Bagheria, Sicily	1985–1994	49.4	16.2–72.9	(Salemi et al. 2000)
Italy	Caltanissetta, Sicily	2002	165.8	158.5–173.1	(Grimaldi et al. 2007)
Italy	Catania, Sicily	2000–2005	127.1	115.1–140.4	(Nicoletti et al. 2011)
Italy	Ferrera	2004	120.93	110.05–134.23	(Granieri et al. 2007)
Italy	Frosinone, Lazio	2007	95	86.6–104.0	(Millefiorini et al. 2010)
Italy	Genoa	1997	94	88–100	(Solaro et al. 2005)
Italy	Linguaglossa, Sicily	1991–2001	203	107–352	(Nicoletti et al. 2005)
Italy	Salerno	2005	71.62	62.03–82.30	(Iuliano and Napoletano 2008)
Italy	Sardinia	1997	144.4		(Pugliatti et al. 2001b)
Italy	Sardinia		(e)		(Sotgiu et al. 2004)
Italy	Sardinia	1997	(e)		(Pugliatti et al. 2009)

Table 3.1 MS prevalence publications—studies in Italy

Publications reporting studies on MS prevalence rates in Italy since 2000 not included in Rosati (2001) or Pugliatti et al (2002).

Notes for Table 3.1, Table 3.2 and Table 3.3:

Study groups included all cases within the given location, except where indicated in parentheses, and with case inclusion criteria as detailed in the relevant reference. Crude prevalence is given where reported, either a single value where one is given for the whole study group, or a range of values where values are only reported for sub groups within the study. Where reported in the studies, 95% confidence intervals (CI) are given for single prevalence values.

- (a) Only standardised prevalence is reported which is given in the table instead of crude prevalence—see the reference for details of the standard population used.

Studies with no prevalence given in the table report:

- (b) incidence rates;
(c) ratios of rates;
(d) other data or rates not based on population numbers;

or

- (e) discuss prevalence while focussing on other factors.

Country	Location (Study Group)	Study Years	Prevalence /10 ⁵	95% CI /10 ⁵	Reference
Austria	National	1998–1999	98.5		(Baumhackl et al. 2002)
Bosnia & Herzegovina	Western Herzegovina	1994–2003	27	20–34	(Klupka-Sarić et al. 2007)
Croatia	Čabar	1948–2004	205.7		(Perković et al. 2010)
Croatia	Varaždin	1991–1997	24.04–30.19		(Dobec-Meić et al. 2008)
Croatia & Slovenia	Gorski kotar & Kočevje	1999	151.9	123.2–187.4	(Peterlin et al. 2006)
Denmark	National	1950–1996	122.5		(Stenager et al. 2005)
Denmark	National	1950–2005	173.3	169.9–176.7	(Bentzen et al. 2010)
Denmark	National (relatives of cases)	1968–1997	(c)		(Nielsen et al. 2005)
Faroe Islands	National (one family)		(e)		(Binzer et al. 2010)
Finland	Central Finland	1979–2000	105	93–117	(Sarasoja et al. 2004)
Finland	Northern Ostrobothnia	1992–2007	103	93–113	(Krökki et al. 2011)
Finland	West & South Finland	1979–1993	(a) 93–188		(Sumelahti et al. 2001)
France	Haute-Garonne	2005	110		(Sagnes-Raffy et al. 2010)
France	National (farmers)	2003	(a) 65	62.5–67.5	(Vukusic et al. 2007)
France	National	2003–2004	(a) 94.7	94.3–95.1	(Fromont et al. 2010)
Greece	Western Greece	1984–2006	119.61		(Papathanasopoulos et al. 2008)
Hungary	Csongrád	1997–1999	62		(Bencsik et al. 2001)
Ireland	Donegal & Wexford	2001	(a) 120.7–184.6		(McGuigan et al. 2004)
Ireland	Donegal, Wexford & Dublin	2007	(a) 123.2–257.8		(Loneragan et al. 2011)
Malta	National	1999	13.2		(Dean et al. 2002)
Norway	Nordland	1970–1999	105.6	93.3–119.5	(Alstadhaug et al. 2005)
Norway	Nord-Trøndelag	2000	163.6	142.2–187.5	(Dahl et al. 2004)
Norway	Northern Norway	1983–1993	73	62.3–85.1	(Grønlie et al. 2000)

[Table continued on next page]

Table 3.2 MS prevalence publications—studies in Europe, excluding Italy

Publications reporting studies on MS prevalence rates in Europe, excluding Italy, since 2000 not included in Rosati (2001) or Pugliatti et al (2002).

Notes: See Notes given for Table 3.1.

Country	Location (Study Group)	Study Years	Prevalence /10 ⁵	95% CI /10 ⁵	Reference
Norway	Oslo (minority races)	2005	148	138–158	(Smestad et al. 2008)
Norway	Northern Norway (Sami)		30		(Harbo et al. 2007)
Portugal	Santarém	1994–1999	46.3	29.5–63.2	(De Sá et al. 2006)
Romania	Mures	2006	26.1		(Bălaşa et al. 2007)
Spain	Alcoi	1986–1997	41.28	31–53.6	(Mallada-Frechín et al. 2000)
Spain	Bajo Aragón	1994–2003	75	52–97	(Modrego and Pina 2003)
Spain	Las Palmas, Canary Islands	1998–2002	77.5	59.7–98.9	(Aladro et al. 2005)
Spain	Menorca	1987–1996	66.8	50.3–91.6	(Casquero et al. 2001)
Spain	Santiago de Compostela, Galicia	1998–2003	78.7	60.4–97.0	(Ares et al. 2007)
Spain	Southern Spain (Gypsies)	2002	52.9	24–82	(Fernández et al. 2008)
Sweden	National	2008	(c)		(Salzer et al. 2010)
Sweden	National	2008	188.9	186.1–191.7	(Ahlgren et al. 2011)
Sweden	Värmland	2002	170.07	154.5–185.5	(Boström et al. 2009)
Sweden	Västerbotten	1990	125	112–140	(Sundström et al. 2001)
Sweden	Västerbotten	1997	154	139–170	(Sundström et al. 2003)
Turkey	Maltepe, Istanbul	2002–2003	101.4		(Börü et al. 2006)
United Kingdom	Devon	2001	118	106–129	(Fox et al. 2004)
United Kingdom	England (hospital admissions)	1998–2005	16.4	16.2–16.5	(Ramagopalan et al. 2011)
United Kingdom	Glasgow	1989–1998	145	127–163	(Murray et al. 2004)
United Kingdom	North east Northern Ireland	2004	230.6	207.0–255.4	(Gray et al. 2008)
United Kingdom	West Scotland	1998–2005	(c)		(Bayes et al. 2010)
United Kingdom	South east Wales	1985–2005	146	135.0–158.0	(Hirst et al. 2009)
Yugoslavia	Belgrade	1985–1996	(a) 41.5	38.5–44.7	(Pekmezovic et al. 2001)

Table 3.2 MS prevalence publications—studies in Europe, excluding Italy [Continued]

Country	Location (Study Group)	Study Years	Prevalence /10 ⁵	95% CI /10 ⁵	Reference
<i>Africa</i>					
Kenya	Private clinic (neurology patients)	1989–1999	(d)		(Kioy 2001)
South Africa	KwaZulu, Natal	1980's–2005	0.22–25.63		(Bhigjee et al. 2007)
<i>Asia</i>					
Japan	Asahikawa	2002	10.18	7.17–14.03	(Itoh et al. 2003)
Japan	National	2004	7.7	7.1–8.4	(Osoegawa et al. 2009)
Japan	Tokachi	2001	8.57	5.82–12.17	(Houzen et al. 2003)
Japan	Tokachi	2006	13.1	9.6–17.4	(Houzen et al. 2008)
Korea	National	2000–2005	3.6	3.2–4.0	(Kim et al. 2010b)
Russia	Amur	2005	30.2		(Karnaukh 2009)
Russia	Bashkortostan		(c)		(Bakhtiarova and Magzhanov 2006)
Russia	Novosibirsk	1970–2003	54.4		(Malkova et al. 2006)
Russia	Tyumen, Khanty-Mansi & Yamal-Nenets	2005	22.4		(Sivertseva et al. 2006)
Russia	Volgograd	1996–2000	(a) 31.9		(Dokuchaeva and Boiko 2006)
Russia	Yamal (Komi-Zyrians)	2005–2008	16.2		(Sivertseva et al. 2010)
Taiwan	National (Chinese)	1985–1999	1.9		(Tsai et al. 2004)
<i>Australasia</i>					
Australia	Capital Territory	1996	(a) 56.7	43.1–74.1	(Simmons et al. 2001)
Australia	Newcastle	1961–1996	59.1	46.3–73.2	(Barnett et al. 2003)
Australia	New South Wales, Queensland, South Australia & Western Australia (UK & Irish immigrants)	1947–1981	28.92–41.78		(McLeod et al. 2011)

[Table continued on next page]

Table 3.3 MS prevalence publications—studies in non-European countries

Publications reporting studies on MS prevalence rates in non-European countries since 2000 not included in Rosati (2001) or Pugliatti et al (2002).

Notes: See Notes given for Table 3.1.

Country	Location (Study Group)	Study Years	Prevalence /10 ⁵	95% CI /10 ⁵	Reference
New Zealand	National	2006	72.4		(Taylor et al. 2010b)
<i>Middle East</i>					
Egypt	Various (neurology patients)		(d)		(Hashem et al. 2010)
Iran	Isfahan	2004–2005	35.5	33.6–37.3	(Etemadifar et al. 2006)
Iran	Isfahan	2003–2006	43.8		(Saadatnia et al. 2007)
Iran	Isfahan	2003–2007	(d)		(Maghzi et al. 2010)
Iran	Khorasan	2009	5.3–12.9		(Ghandehari et al. 2010)
Iran	Tehran	1989–2009	55.98	54.75–57.21	(Elhami et al. 2011)
Israel	Jerusalem	1995	(a) 19.2–64.3		(Karni et al. 2003)
Israel	National	2000	5.4–67.9		(Alter et al. 2006)
Sudan	National	1998–2008	(e)		(Idris et al. 2009)
UAE	Dubai	2000–2007	54.77	46.99–62.55	(Inshasi and Thakre 2011)
<i>North America</i>					
Canada	Alberta	2005	(a) 335.0	328.5–341.5	(Svenson et al. 2007)
Canada	Alberta	1990–2004	(a) 357.6	351.0–364.2	(Warren et al. 2008)
Canada	Manitoba	1984–2006	278		(Marrie et al. 2010)
Canada	Newfoundland & Labrador	1958–2001	94.4	90.2–98.7	(Sloka et al. 2005a)
Canada	Newfoundland & Labrador	2001	(e)		(Sloka et al. 2005b)
Canada	Saskatoon, Saskatchewan	1970–2005	298.3	274.7–323.6	(Hader and Yee 2007)
Canada	Toronto	1993–2004	(e)		(Kennedy et al. 2006)
French West Indies	National (West Indian)	1998–1999	14.8	11.9–17.7	(Cabre et al. 2005)
USA	Alaska (military veterans)	1960–1994	22.35	8.97–46.03	(Wallin et al. 2009)
USA	Jefferson County, Missouri	1998–2002	105	91–121	(Turabelidze et al. 2008)
USA	National (military veterans)	1960–1994	(d)		(Wallin et al. 2004)
USA	Olmsted, Minnesota	2000	176.7		(Mayr et al. 2003)

[Table continued on next page]

Table 3.3 MS prevalence publications—studies in non-European countries [Continued]

Country	Location (Study Group)	Study Years	Prevalence /10 ⁵	95% CI /10 ⁵	Reference
USA	Texas	1998–2000	42.8	36.8–49.5	(Williamson et al. 2007)
<i>South America</i>					
Argentina	Metropolitan Buenos Aires	1996	14–19.8		(Cristiano et al. 2009)
Argentina	Pategonia	2002	17.2		(Melcon et al. 2008)
Brazil	Santos, São Paulo	2005	15.54		(Fragoso and Peres 2007)
Brazil	São Paulo	1997	15		(Callegaro et al. 2001)
Chile	National	2002–2006	(b)		(Díaz et al. 2011)
Colombia	Antioquia, Caldas, Santander,	1995–2000	1.48–4.98		(Sánchez et al. 2000)
Colombia	Bogotá	2002	4.41	3.9–4.9	(Toro et al. 2007)
Ecuador	Cuenca, Guayaquil & Quito	2006	0.75–5.05		(Abad et al. 2010)
Panama	National	2000–2005	5.24	4.49–6.07	(Gracia et al. 2009)

Table 3.3 MS prevalence publications—studies in non-European countries [Continued]

Country	Reference
Europe	(Pugliatti et al. 2006)
Middle East	(Al-Hashel et al. 2008); (Benamer et al. 2009)
Northern Hemisphere	(Willer et al. 2004)
Latin America	(Corona and Román 2006); (Risco et al. 2011)
South America	(Cristiano et al. 2008)
World	(Simpson et al. 2011); (WHO and MSIF 2008)

Table 3.4 MS prevalence publications—studies covering several countries

Publications reporting studies on MS prevalence rates which cover several countries since 2000 not included in Rosati (2001) or Pugliatti et al (2002).

Table 3.3 lists studies reported for non-European countries. There is still a distinct lack of studies based in Africa and most of southern Asia, though there are a number of new studies based in the Middle East and South America which help to fill previous gaps in the global prevalence map. A number of multi-national studies have also been reported which collate results from other studies. These are listed in Table 3.4

The most comprehensive estimates for global variations in MS prevalence rates are as reported in the WHO MS Atlas (WHO and MSIF 2008) which provides data as a collation of responses received from WHO member states. Though only 92 out of 208 members provided actual estimates of MS prevalence rates, Figure 3.2 shows the extent of data now collated (MSIF 2008).

The pattern presented by the map of prevalence rates in Figure 3.2 is very similar to that presented by the map of death rates in Figure 3.1. Europe, Canada, USA and Australasia have the highest rates whilst southern Asia and central America have the lowest. On this evidence it is difficult to distinguish between any possible latitude effect and an effect due to genetic ancestry. What is more likely, however, is that both effects are combining and where there are ethnic mixtures the genetic effect is stronger than any latitude effect. If this is the case then areas which extend across a broad latitude range, but which exhibit a reasonably isolated genetic base in the population, should see latitude effects dominate. This is likely to be what was seen in Alter et al (1973): when the North America data was looked at in isolation it showed a strong latitude effect; when all the data available was combined it was difficult to find a simple relationship between prevalence and latitude. Therefore, in order to study any latitude effect, any ethnic genetic variability needs to be taken into account either by only including one ethnicity in the study, or by stratification of the data.

The study by Wallin et al (2004) also looked at a possible latitude gradient within the USA by studying locations of residence of US war veterans prior to entry in the service. After stratification of the data by gender and ethnicity—white, black and 'other' (Indian and Asian)—they found evidence of a gradient in every group. What is just as notable, however, is their comment regarding a possible genetic admixture within their cohort of cases which could confound the results. The authors point out that most of their cases were born before 1970 and it is only within the last few decades that inter-racial marriages have become more common, with a 7 fold increase

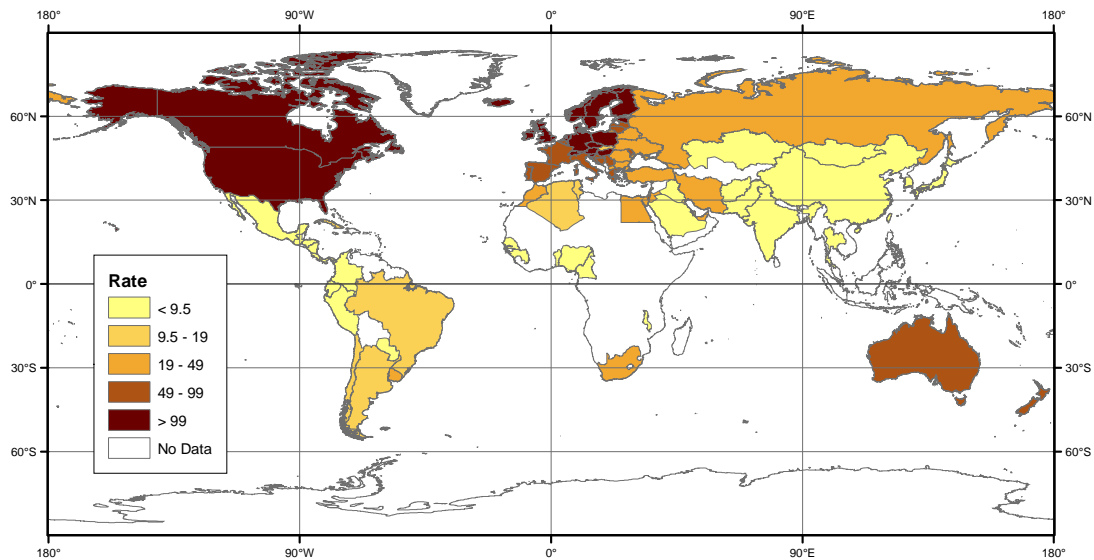


Figure 3.2 Age-adjusted prevalence rates for MS by WHO member state
 MS Prevalence rates by WHO member state per 100,000 population for 2005–7, age adjusted by the WHO global standard population. (MSIF 2008).

between 1960 and 1990. As such, they believe that any confounding effects of ethnicity have been taken into account. For this study, that is probably true. However, it suggests that future studies may not find it so easy to separate out ethnicity effects from effects due purely to other factors such as latitude. Instead of simply stratifying by broad ethnic categories, it will become increasingly important to stratify by underlying genetic code at the genotype and allele level, as in the study by Lonergan et al (2011) which looks at MS in Ireland. In that study, latitude is considered alongside vitamin D levels, but it is the frequency of the HLA DRB1*1501 allele (as discussed in section 2.2.3.2) which is seen to be key when combined with vitamin D.

This is also the approach taken by Simpson et al (2011) who performed a meta-analysis study based on 650 prevalence estimates from 321 other studies. A global positive prevalence–latitude gradient was confirmed, a relationship which strengthened when the prevalence year was taken into account, though this gradient was not uniform globally and there were exceptions. Most apparent was that there did not appear to be a gradient for nations with populations of non-European ancestry. So although MS should no longer be considered solely a European disease, it would appear that the prevalence–latitude gradient effect might be a European genetic phenomenon. To further complicate the situation, two of the European regions, Italy and "Scandinavia and North Atlantic", exhibited negative gradients in keeping with previous studies

(Rosati 2001). North America and Australasia had almost identical positive, exponential trends in gradients across the full range of latitudes (16–55° N or S). When the HLA DRB1 allele frequency was taken into account for the Italian region, the gradient reversed to give a positive trend similar to the rest of Europe. This suggests that the inverse gradient in the Italian region is due entirely to the unique distribution of the HLA DRB1 alleles in this region, as described in section 2.2.3.2. A lack of data on allele frequency in the Scandinavian region meant that a similar analysis could not be performed to establish whether or not the mix of Sami ancestry could account for the negative gradient. However, the authors refer to a study by Kampman and Brustad (2008) who note that the Scandinavian diet is much higher in vitamin D than the diet of the rest of Europe, and that winter vitamin D deficiency was actually more common in southern Europe than in Scandinavia. This higher dietary intake of vitamin D could contribute to the inverse gradient observed in this region.

A number of studies have looked at MS rates in Australia and New Zealand, countries which span roughly 33° and 13° of latitude, respectively. Although both countries have indigenous populations, Australian Aborigines and New Zealand Māori being the main populations, the non-indigenous populations are mainly white and of UK origin. In addition, as with the Sami, it has been observed that there are very few cases of MS in these indigenous populations. Both countries should provide insight into the MS prevalence–latitude relationship as illustrated by the study of Miller et al (1990). (Although this study was covered by the review of Rosati (2001), it is useful to include a summary here.) In that study, data from six regions of Australia and three from New Zealand were combined. The Figure of Miller et al (1990) is reproduced in Figure 3.3(a), with several additions, using data from Table 1 of Miller et al. A strong prevalence–latitude gradient was observed, though variations were also seen which could be attributed to differences in case ascertainment between the underlying studies, and possibly to genetic mixing in the north of New Zealand (Waikato). The largest Māori population is in the north so if there is any significant mixing between Māori and the non-Māori population then effects are most likely to be in Waikato. Also shown in Figure 3.3(a), which was not in the original study of Miller et al, is an exponential fit to the data. Interestingly, this appears to be somewhat better than the original linear fit. This is similar to the finding of Alter et al (1973), described earlier, where an exponential fit to the North America data was better than any of the other simple fits

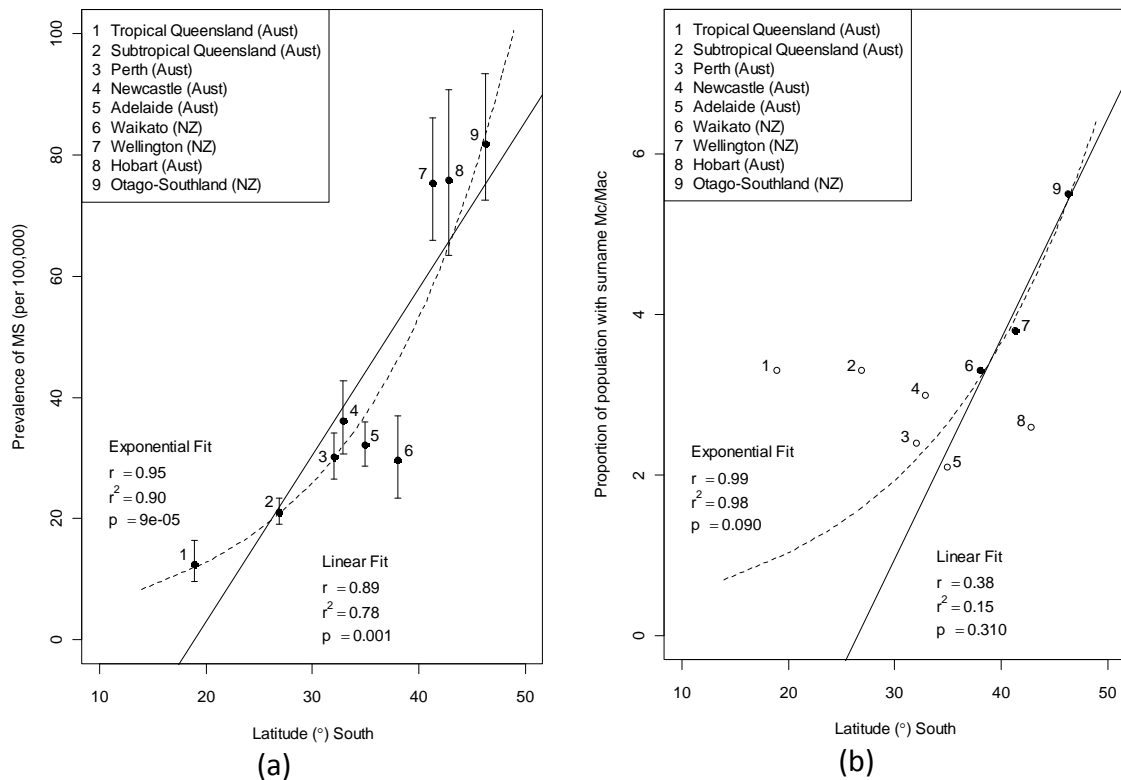


Figure 3.3 MS in Australia and New Zealand

(a) Prevalence of MS in Australia and New Zealand by latitude, after the Figure in Miller et al (1990). Data is from Table 1 of Miller et al. Confidence intervals are from that table; they were not shown in the original Figure. Least squares regression fits: Solid line: linear fit from the original Figure; Dashed line: exponential fit determined for this figure. **(b)** Proportion of population with surname "Mc" or "Mac" in Australia and New Zealand by latitude. Data is from Table 2 of Miller et al. Locations: Open circles: Australia; solid circles: New Zealand. Least squares regression fits to New Zealand data determined for this figure: Solid line: linear fit; Dashed line: exponential fit.

which were investigated. The use of exponential (log) fits will be discussed in more detail in Chapter 4.

Although both Australia and New Zealand were settled by immigrants from all over the UK, not all areas were settled equally. Given that the prevalence rate of MS in Scotland tends to be much higher than that in England, as discussed in section 2.2.3.2, any variations in the numbers of Scots settling in an area could give rise to a founder effect, where variations in the observed prevalence of MS may be due to genetic inheritance in addition to geographical location. In particular, more Scots tended to settle in the south of New Zealand. In addition to looking at prevalence of MS, Miller et al examined a crude indicator of the degree of Scottish ancestry within the areas studied by counting the number of surnames beginning with "Mc" or "Mac" in the relevant telephone directories. The proportions obtained were 2.1–3.3% for Australia and 3.3–5.5% for New Zealand, as illustrated in Figure 3.3(b). (The regression fits

shown for the New Zealand data are intended as a guide to the eye rather than any suggestion of functional form within the data.) Although Miller et al found no significant correlation with latitude when both the Australia and New Zealand data were considered, and there is no significant correlation when considering the New Zealand data on its own, there is a strong suggestion that the degree of Scottish ancestry in the New Zealand population is not independent of latitude within New Zealand. As such, any genetic factors which may affect the prevalence of MS should be considered as possible co-variables with latitude when considering the prevalence of MS with latitude within New Zealand.

Although several prevalence studies have been carried out in Australia since Rosati (2001), all have looked at local prevalence rates. The most recent extensive study within New Zealand was carried out nationally and attempted to include all MS cases resident in New Zealand at the time of the 2006 census. This study is described further in the next section.

3.3 The New Zealand National MS Prevalence Study

The New Zealand National MS Prevalence Study (NZMSPS) was carried out alongside, and subsequent to, the 2006 New Zealand population census by a team of experts in the fields of neurology and health data analysis (Taylor et al. 2010b). The team established a database of MS cases resident in New Zealand on the date of the census and collected a wealth of data pertinent to the study, including details of gender, MS phenotype, self-declared ethnicity and family history of MS. The data also included a *residence calendar*, a record of past residence locations and an estimation of the time spent in the sun at each location. Further details of the NZMSPS project are given in section 3.3.1 with initial results outlined in 3.3.2. Details of the residence calendar are then given in section 3.3.3.

3.3.1 The NZMSPS Project

The NZMSPS was jointly funded by the NZ Health Research Council and the National MS Society of NZ with the specific aim of determining the national prevalence of MS in New Zealand on census day, 7 March 2006. The investigators are given in Table 3.5 and comprise a team of experts in neurology, health science, and GIS.

From July 2005, the NZMSPS built up a database of MS cases from multiple sources, including databases from MS societies, hospitals, and neurologists. Information was also gathered from NZ government health information statistics services, MS care providers, and through direct advertising. All patients (cases) were notified to the team by a notifier who provided the case with a unique identifying code comprising date of birth, gender and initials. If the code was unique when received by the study centre then the notifier was asked to invite the patient to participate in the project. Informed consent for participation in the project and access to medical records was sought, and confidentiality was ensured at all times. All unique cases had their diagnosis of definite multiple sclerosis (DMS) confirmed by a neurologist after reviewing the case notes and, in many circumstances, a face-to-face interview, and all were confirmed as being resident in New Zealand on census day. Unique cases confirmed with DMS and resident on census day formed the cohort of participating cases; these cases were sent the NZMSPS questionnaire. Those who failed to return the questionnaire were contacted, either by their notifier, or by the study centre, according to consent given. Where required, assistance in completing the questionnaire was provided, either directly by research staff, or by using the help of relatives.

The questionnaire which cases were asked to complete consisted of three sections, the main questionnaire, a residence calendar and a job calendar. The residence calendar will be looked at in more detail in section 3.3.3 since it forms the data upon which this dissertation is based; the job calendar was included to provide information on whether jobs which the case had held prior to onset of MS were mainly indoor or outdoor jobs.

Investigator	Role	Affiliations
A/Prof Bruce Taylor	Consultant Neurologist	Christchurch School of Medicine and Health Sciences, University of Otago (Now University of Tasmania, Australia)
Dr Clive Sabel	GIS and Health Science	University of Canterbury (Now University of Exeter, UK)
A/Prof Ann Richardson	Epidemiologist	Christchurch School of Medicine and Health Sciences, University of Otago
Dr Debbie Mason	Consultant Neurologist	Christchurch Public Hospital, Canterbury DHB Christchurch School of Medicine & Health Sciences, University of Otago
Dr David Abernethy	Neurologist	Capital and Coast HB
Dr Ernie Willoughby	Neurologist	Auckland City Hospital, Auckland DHB Auckland University School of Medicine

Table 3.5 NZMSPS project investigators

NZMSPS project investigators giving their roles and affiliations at the time of the study.

1. What age group are you in?
2. a) Where did you live at the time of the 2006 census (March 2006)? b) Where were you born ?
3. a) What year were you first diagnosed with MS? b) Where were you living?
4. a) What year did you first have symptoms of MS? b) Where were you living?
5. Which ethnic group do you belong to?
6. What is your ancestral group? This refers to your ancestral origins (heritage) and may be different from your ethnic group.
7. As best you can, please give your grandparents' ancestry Maternal Grandparents (your mother's parents) Paternal Grandparents (your father's parents)
8. What is your highest secondary school qualification?
9. Apart from secondary school qualifications, do you have another completed qualification? <i>DON'T count qualifications that take less than 3 months of full-time study to get</i>
10. Print your highest qualification, and the main subject, for example: Qualification <i>TRADE CERTIFICATE</i> ; Subject <i>ELECTRICAL ENGINEERING</i>
11. In the last 7 days, which of these did you do? [Options relate to type of work, paid or unpaid]
12. In that job, what was your occupation, for example: <i>PRIMARY SCHOOL TEACHER, CLOTHING MACHINIST, MOTEL MANAGER, RECEPTIONIST?</i>
13. How many hours, to the nearest hour, in all your jobs (for profit or unpaid in a family business/farm) do you usually work each week?
14. Has your job (occupation or hours of work) ever changed as a result of MS?
15. Please specify how your job has changed for the most recent time as a result of MS
16. What was the occupation you were originally trained for? For example: <i>PRIMARY SCHOOL TEACHER, CLOTHING MACHINIST, MOTEL MANAGER, RECEPTIONIST?</i>
17. Tick as many boxes as you need to show all the ways you yourself got income in the 12 months ending today
18. From all the sources of income you marked in question 17, what was the total income that you yourself got before tax or anything was taken out of it, in the 12 months ending 31 March 2006?
19. Which of these statements is true about your legal marital/civil union status?
20. Where do you live at present? [Options relate to type of dwelling]
21. Where did you live 5 years ago?
22. At present, do you receive services from any of the following? [Options relate to health care services]
23. Does anyone in your immediate family (parents, siblings, half-siblings) have, or has anyone had multiple sclerosis?

Table 3.6 A summary of Section 1 of the NZMSPS questionnaire

A summary of the questions which were asked in the NZMSPS questionnaire. The full questionnaire is given in Appendix A.

In addition to these three sections, the questionnaire also contained two consent forms which would be held separately from the rest of the questionnaire to ensure the confidentiality of the case. Questions included in Section 1 of the questionnaire are given in Table 3.6, in summary form; the full questionnaire, with consent forms, is given in Appendix A. Where appropriate, responses were sought in categories rather than free text. The questions were grouped into several broad categories, including ethnicity and the case's own and family history of MS. Part of the MS diagnosis procedure would also confirm which MS phenotype the case had.

The next section continues with a summary of initial results from the NZMSPS project.

3.3.2 Prevalence of MS in New Zealand by Region

The initial NZMSPS results (Taylor et al. 2010b) looked at prevalence of MS within New Zealand on census day 7 March 2006. The crude prevalence of MS was 72.4 per 100,000, which increased slightly to 73.1 per 100,000 when age-standardized to the European standard. An in depth look at all of the results from Taylor et al is not suitable here, though some of those relating to latitude and MS phenotype are worth noting for later comparison.

Table 3.7 gives a summary of the MS phenotype data from Table 1 of Taylor et al. It can be seen that the RR/SPMS phenotype constitutes 84% of the cases, a figure which is consistent with the general trend mentioned in Chapter 2. It is also noticeable that the female to male ratio is significantly higher for the RR/SPMS phenotype than for the PPMS phenotype.

The effect of latitude was examined by stratifying the data by region of residence. For these analyses, the six regions considered were artificially constructed from aggregates of the 16 New Zealand regions as illustrated in Figure 3.4. The representative latitude

MS phenotype	Number of cases	Male:Female ratio
PPMS	458 (16%)	1:1.5
RRMS	1541 (53%)	1:3.6
SPMS	918 (31%)	1:4.5
RR/SPMS	2459 (84%)	1:3.9
All	2917	1:3.0

Table 3.7 Summary findings of MS phenotype from the NZMSPS

Data for RRMS, SPMS, PPMS and All from Table 1 of Taylor et al (2010b). Data for RR/SPMS is derived here. MS phenotypes are as described in section 2.2.1.

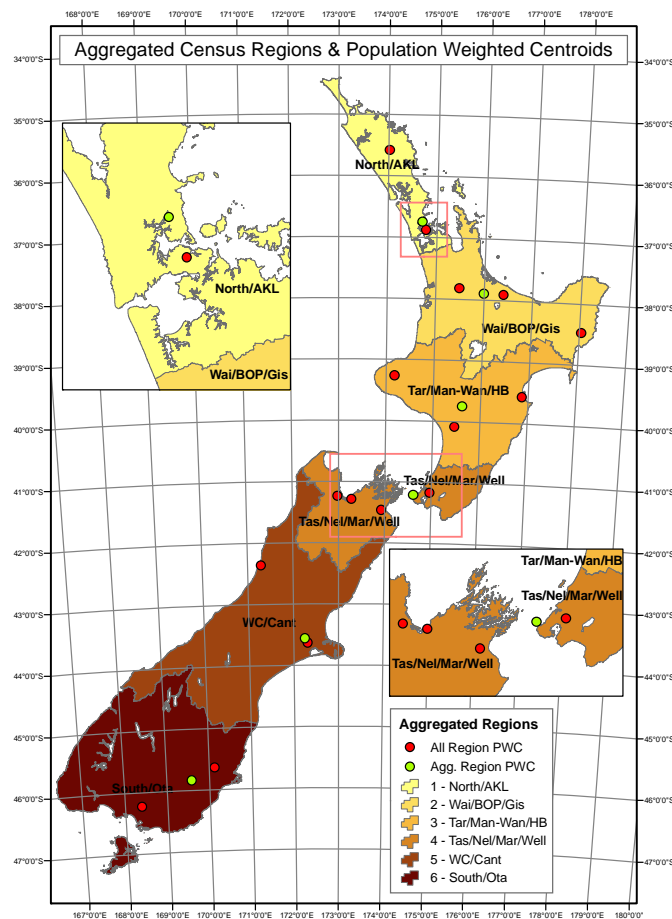


Figure 3.4 Aggregated regions in New Zealand

From Taylor et al (2010b). The aggregated census regions used in the study. Population weighted centroids (PWC) for the New Zealand statistical regions and the aggregated census regions are shown.

for each aggregated region was taken to be the population weighted centroid of the aggregated region. A visual examination of the map in Figure 3.4 indicates that the latitudes of these centroids are close to the centres of the aggregated regions for all regions except for Region 1, the most northerly region. Region 1 includes the city of Auckland in the south of the region which has one quarter of the total New Zealand population; this is sufficient to influence the centroid and bring it south, closer to that of Region 2. Figure 3.5 gives prevalence of MS against latitude by gender and phenotype. Distinct prevalence–latitude gradients are found, which vary by both gender and phenotype. The prevalence–latitude gradients are greatest for the RR/SPMS phenotype.

Table 3.7 gives male–female ratios for the different phenotypes. As well as varying by phenotype, it would appear from Figure 3.5 that the ratio also varies with latitude within the different phenotypes, though this is not discussed in the paper.

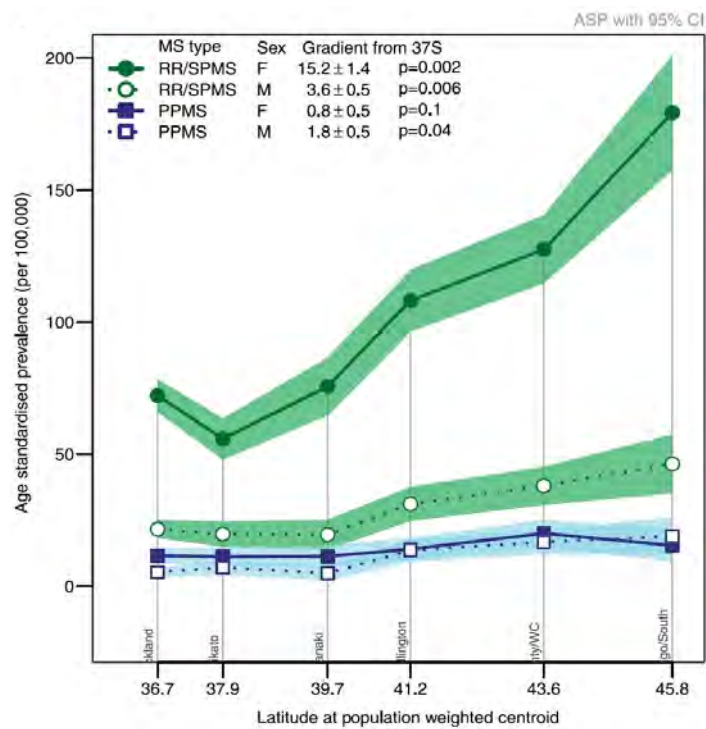


Figure 3.5 Prevalence of MS in New Zealand 2006

From Taylor et al (2010b). Prevalence of MS by gender and MS phenotype. The gradients are from linear models to the data from 37°S.

The main finding from these initial results, which is of relevance to this study, is the very strong prevalence–latitude gradient which varies by gender and MS phenotype. The results also suggest that the prevalence rates tend to increase steadily with latitude rather than being subject to incremental change.

The study by Taylor et al only looked at prevalence on census day 2006. For the study reported in this dissertation, the prevalence rates of MS at various points in the lifetime of the cases will be looked at based on data from the residence calendar provided by each case. Not only can certain ages be considered, for example, Age 0–4, but also periods of life relative to the onset of MS. The residence calendar is discussed in more detail in the next section.

3.3.3 The Residence Calendar

Each MS case identified by the NZMSP study was asked to provide a duration and location of residence for each place of residence during their lifetime. In addition, an indication of the time spent in the sun, in summer and in winter, was also requested in

terms of hours per week spent in the sun. Section 2 of the questionnaire which was used is illustrated in Figure 3.6 and in full as part of the questionnaire in Appendix A. Although the data collected from Section 1 of the questionnaire would provide an indication of prevalence on Census day, as has already been described in section 3.3.2, the residence calendar provides data which can be used to investigate the spatial variations in prevalence at various times in the case history.

The data collected included free text locations of residence along with the dates at those residences. The NZMSPS examined each location provided and converted the locations into latitude and longitude co-ordinates prior to the data being included in this study. Given the date of birth of the case, and the year of onset of MS, this

NZMS Prevalence Study				
<i>Office use only</i>	From	To	Where you've lived	
	(Month and year)	(Month and year)	Suburb and city, or nearest town and district	
Period	1	2	3	4
	Until diagnosis	Until diagnosis	Suburb / Town	City / District (or country)
example	6 / 1962	12 / 1975	Highbury	Wellington
R1	Conception	__ / __		
R2	__ / __	__ / __		
R3	__ / __	__ / __		
R4	/	/		

Town and District (or country)	Sunlight Exposure										Sunburn				
	Average hours per week spent in the sun										Number of times sunburnt				
	Place tick ✓ in appropriate box										Place tick ✓ in appropriate box				
	Winter					Summer					0	1	2-5	6-10	19
	<5	5-9	10-19	20-29	30+	<5	5-9	10-19	20-29	30+					
Wellington															

Figure 3.6 The Residence Calendar which MS cases were asked to fill out
 Details requested included dates and location for each residence location and estimated exposure to sunlight.

information was sufficient to establish the latitude of residence of the case at any month of life or any year relative to onset of MS. It would also be possible to examine the effects of migration on the prevalence rates of MS.

3.3.4 Limitations of the NZMSPS Data

The main limitation of this study is that, despite having a significant amount of data to analyse, there is no control data. Although this is not critical when looking at age-standardised prevalence data, Chapter 6 will show that it is significant for looking at UV data. This limitation reduces the analyses available to those which examine the variations within the MS case data; analyses which try to establish how the MS data varies from that of the general population are not possible.

The residence calendar asked for estimated summer and winter exposure to sunlight for each residence location. Chapter 6 will show that further information on time of day of that exposure, or, for example, week-day and weekend patterns, would have been very useful since the variations in the strength of sunlight are considerable—for mid summer in New Zealand, 2–3 minutes of exposure to the sun around midday is just as significant for vitamin D synthesis as an hour or so close to dusk, though angle of exposure is equally important. It could be argued that people would either not know or not be able to remember time of day details many years on, but these details may be what is remembered and what the weekly totals reported in the residence calendar are then based on. Uncertainties in the data due to these factors are likely to outweigh any misclassification introduced by recall bias, which is difficult to avoid in any questionnaire based data. Recall bias here, could be introduced either from an individual's lack of memory due to passing years or ill-health, or from a conscious or subconscious desire to bias the data provided. In the latter case, an individual could be biased towards finding a cause to blame for their MS, or they could be trying to avoid blame for perceived behaviour on their part. How this affects the data provided also depends on what the individual's understanding is of any link between MS and sunlight exposure. However, since there are no control cases for the exposure data, it is not possible to attempt to quantify any differential bias which may have been introduced.

Having established that the MS data which is available for analysis includes residence locations, it is important to know what the equivalent general population data is to enable prevalence to be calculated. Specific details of the data required and numerical

methods are given later in Chapter 4, however, it is necessary to know here that, in the absence of control data, the underlying census population data for the period of interest is required for normalisation and age standardisation of the prevalence rates. In addition, for the periods of interest here, the census data differs in terms of the census areas used to underpin the data. Drawing together different geographic data into a single structure requires a technique such as areal interpolation which is discussed in the next section.

3.4 Areal Interpolation and the MAUP

Areal interpolation and the modifiable areal unit problem (MAUP) overlap in certain respects, but there are also aspects of both which are independent of the other. The MAUP is concerned with implications of the choice of aggregation technique used to aggregate spatial data; areal interpolation is one technique which can be used to achieve the chosen aggregation. It is useful to start with a description of the MAUP.

3.4.1 The Modifiable Areal Unit Problem

The modifiable areal unit problem (MAUP) essentially states that "analytical results for the same data in the same study area can be different ... if aggregated in different ways" (Flowerdew et al. 2008). An in depth investigation of the MAUP is beyond the remit of this study, however, given that the study does require aggregation of data at some level, it would be prudent to consider implications of the aggregation options.

Fundamentally, in order to investigate any relationship between prevalence of MS and latitude, the data available has to be aggregated into a spatial structure (geography) which can provide a representative latitude. For the initial NZMSPS results presented by Taylor et al (2010b), the geography used was an aggregation of the administrative regions within New Zealand into six composite regions. If census data is required here for further analysis of the results then a similar choice of regions would be an option. However, this study will investigate the prevalence of MS at various life periods of the MS cases and, as such, historical census data will be required, as will be discussed in Chapter 4. The complication with this is that the geography for the census changes with time; the geography for the 2006 census was different from that for 1986 and different again from that for 1956. So the choice then becomes: should the results simply be reported for the geography of the census data concerned, or should the data

be aggregated into some independent geography, in which case, what should that geography be?

The choice made here is to use an independent geography, and since the interest here is only with latitude rather than with any component of longitude, one option for the geography is to simply divide New Zealand by latitude. This has several advantages:

- It removes ties with the underlying, artificial administrative areas.
- The representative latitudes should be reasonably well spaced and, given the shape of New Zealand, the new areas should be of similar sizes.
- The number of areas can be changed easily without requiring a different structure, which would enable a basic investigation into any MAUP effect.

It also has two major disadvantages when compared with the region approach:

- Administrative areas tend to be of a similar size in terms of population whereas areas defined by latitude are likely to have populations which vary widely.
- Regional aggregation only requires knowledge of the region concerned whereas latitude based areas require more precise location details.

The most obvious alternative would be to use the 2006 census geography, though this would have none of the advantages listed above, and would still have an equivalent of the second disadvantage since the past residence data available to this study is in terms of location rather than region.

Having established which geography will be used for the data, the next step is to determine how to aggregate the data into that geography. This is often achieved using areal interpolation, which will be outlined next.

3.4.2 Areal Interpolation

A problem which is encountered in many analyses which rely on census population data (though not restricted to such cases) is often referred to as a change of support problem. The 'support' referred to is the geography underlying a given set of data and the problem arises when that geography changes or when several sets of data, based on different geographies, need to be combined. This is not a new problem. However, how the problem is addressed has changed over the years. The conventional approach for combining data from different geographies was to use a representative point such

as an area's centroid. Statistics would be assigned to the centroids of the source areas which would then be interpolated to give statistics for the centroids of the target areas. Although this approach is simple, it is not an ideal solution and the interpolation process may not be trivial. Whether it is an acceptable solution also depends upon a number of factors, not least the availability and the geography of the data.

An alternative approach is to consider the areas of overlap between the areas in the different geographies. This idea was used by a number of studies before being formalised by Goodchild and Lam (1980) and referred to as 'areal interpolation', with the original geographic areas being the 'source' areas and the required areas being the 'target' areas. In principal, the process is as follows, with population as the example for convenience, without any loss of generality:

- The source population is assumed to be uniform within each source area, so each area is assigned its own population density.
- The target geography is compared with the source geography and all areas of overlap between the source and target areas are determined.
- Each overlap area is assigned the population density of the source area which it overlaps—this disaggregation assumes that the population density within each source area is uniform across that area.
- The sizes of the overlap areas are combined with the population densities to give a new population for each overlap area—this reaggregation assumes that the population density within each target area is uniform across that area.
- The overlap areas are aggregated into the target geography with the populations of the constituent overlap areas also being aggregated to give individual populations for the target areas.

Although simple in principle, in practice this process requires software to enable it to be carried out efficiently, unless the source and target areas are almost trivial.

The widespread use of computers and the availability of software for detailed analysis of data has had several important effects—Geographical Information System (GIS) software, which can be used for exploring and analysing geographic data, has become easier to use and more accessible, and the functionality and capabilities available within the GIS have increased enormously. Analyses which would not have been

possible several decades ago can now be considered as standard. Areal interpolation is one such process.

Since population density is not usually uniform, many more sophisticated techniques have been developed to improve the disaggregation and reaggregation processes. Some of these use more advanced interpolation techniques while others include meta data such as land use to estimate variations in the population (Aubrecht et al. 2009; Brindley et al. 2005; Cromley et al. 2009; Flowerdew and Green 1992; Flowerdew et al. 1991; Goodchild et al. 1993; Gotway and Young 2008; Gregory 2002; Gregory and Ell 2005, 2006; Harris and Chen 2005; Langford et al. 1998; Langford 2006; Liu and Liu 2008; Liu et al. 2008; Mennis and Hultgren 2005; Mu and Wang 2008; Mugglin et al. 2000; Murshed 2009; Petrovskii and Li 2003; Reibel 2007; Sadahiro 1999, 2000; Saporito et al. 2007; Su et al. 2010; Tiebei et al. 2007; Wu and Murray 2005; Yuan et al. 1997). Discussion of these techniques is beyond the scope of this dissertation.

The use of areal interpolation for the data used in this study will be discussed further in the next chapter. Meanwhile, another important consideration for the geography of MS and population data, migration, is discussed in the next section.

3.5 Migration

Ancestral migration, leading to the founder effect, was discussed in section 3.2. Equally important for MS and many other diseases, if not more so, is migration within the individual's lifetime. This is considered by studying population migration.

Population migration can confound an otherwise clear relationship (Bentham 1988). As Sabel et al summarise:

"Many studies examining associations between geographical patterns of disease and causal factors assume that current residence in an area can be equated with exposure to conditions that currently (and historically) pertain there (Bentham, 1988). This is important, since the place of residence at the time of diagnosis or death is often adopted by epidemiologists and geographers as the location for further analysis of the disease in question. Yet people move, and hence previous exposure to pathogens will not be included in the study. The problems will be greater for diseases that have a long lag or latency period, allowing plenty of time for mobility of the population. By adopting only the current residential address, not only will an individual's migration history be neglected, but additionally the daily 'activity spaces' of the patient will be ignored." (Sabel et al. 2000, p.1122)

Migration effects are crucial in determining whether or not the age at which an individual is exposed to latitude is significant in determining the risk of MS. If there is no migration then the prevalence rates for a given cohort of cases should not change depending on whether prevalence at onset of MS is considered or prevalence at birth. If there is migration, however, then differences in the rates of prevalence should be apparent. Migration modelling techniques may also be required to establish whether or not any changes in prevalence are due simply to the shift in population or to changes in the risk of MS.

Population migration can be modelled using many techniques from simple matrix equations to complex non-linear integral equations. In basic terms, there are three sub-populations to consider at any one time: those who stay within the area; those who leave; and those who join. Those who leave may be emigrants from the area, or they may die. Those who join may be immigrants to the area, or they may be born. The complexity in migration models depends on how these quantities are determined—are they static proportions based on current populations or are there other factors to take into account? In addition, are there several regional areas with internal migration to consider or does the migration depend on age structure? A simple matrix model, and its application to internal migration, will be discussed further in the next chapter.

3.6 Summary of Geography

The geography of MS is complicated. With all other factors being equal, there is a prevalence–latitude gradient within populations of European ancestry with prevalence increasing with distance from the equator. The correlation appears to be inverse with UV radiation exposure and photosynthesized vitamin D. It is uncertain whether this gradient extends to non-European populations. Factors which confound this gradient include dietary intake of vitamin D and also ethnicity with certain alleles showing a strong correlation with prevalence of MS. Ancestral migration of a population, combined with genetic or ethnicity factors, can give rise to the founder effect. In addition, migration within the lifetime of an individual and maternal migration prior to the individual's birth are confounding factors. Looking at prevalence rates for locations at birth may indicate whether location during childhood and adolescence is more critical in determining risk of MS than location at onset.

Chapter 4

Methodology—Data and Techniques

4.1 Introduction

This study extends the analyses described in section 3.3 to look at different periods in the life of the MS cases. In this chapter, requirements of the age standardisation process and implications in terms of census population data are discussed, along with a description of the GIS and regression techniques employed. A brief discussion on the effects of inter-regional migration is also given.

Standard techniques which are described will be quoted without reference. There are numerous publications which provide details of statistical analysis techniques for epidemiological studies. Two excellent sources are the books "Epidemiology" by Rothman (2002) and "An introduction to medical statistics", Bland (2000).

Before analysis techniques can be discussed, it is important to understand the data which is available for use in the study, so that the required analysis techniques can be determined. Therefore, the next section will outline the data available and the following section will then look at techniques.

4.2 Data Considerations for Different Life Periods

Prevalence of MS is usually quoted as a rate per 100,000. Whether this is crude prevalence or the age-standardised rate, which is described in detail in section 4.3.2, calculation of prevalence requires a knowledge of the population of the area being studied. Consequently, there are two essential prerequisites for determining spatial variations of prevalence: a geography, giving the spatial reference to be used, and the population distribution corresponding to that geography. These will be discussed later in section 4.2.3. First, it is necessary to consider the cohort of MS cases being studied (section 4.2.1), and the period of study (section 4.2.2), to determine which population is required.

4.2.1 Residence Criteria for the Cohorts of MS Cases

This study looks at the prevalence of MS in New Zealand at different periods in the life of the MS cases. This restricts the cases studied to those who were resident in New Zealand at the life periods of interest. By definition, all cases were resident in New

Zealand on census day, 7 March 2006, however, some could have been born in another country and migrated to New Zealand, others could have been born in New Zealand, migrated to another country and then returned to New Zealand. This study considers two cohorts of cases: a 'Migrated' cohort which includes cases who have migrated to or from New Zealand at some point in their lives; and a 'Resident' cohort of cases who have remained entirely resident in New Zealand. For simplicity, for the Resident cohort, the restriction of residence within New Zealand is applied from birth to onset of MS, though the techniques employed could apply the restriction to any period, for example, just at birth. The tighter restriction has the benefit of removing all confounding factors due to migration to or from New Zealand for this cohort. As well as any geographical effects, migration can also confound through lifestyle changes. For example, individuals who have lived in countries which endorse the fortification of foods with vitamin D may have had a higher vitamin D status than could be accounted for by location of residence. (See section 2.2.3 for various discussions on vitamin D.) Some of the analyses carried out in this study were performed on both the Migrated and the Resident cohorts; others which consider the life periods of the cases were only carried out on the Resident cohort.

It should be noted that a third group of 912 cases included in the original cohort of Taylor et al (2010b), described in section 3.3.2, was excluded from this study before any residence criteria were considered, due to a lack of detailed residential history data. Of this 'Missing' group: 707 did not return the questionnaire; 104 declined to participate; 35 died between the census date and the questionnaire data being collected; 32 had incomplete data; 24 were unable to complete the questionnaire due to sickness or disability; and 10 returns were too late for the study. The main issue with reducing the size of the cohort, by omitting those who did not provide data, is whether an unsystematic bias has been introduced: if the cases who did not provide data (for whatever reason) are different from those who did, then the remaining cohort is not representative of all cases. In addition, the residence criterion imposed on the rest of the MS cases, stratified the cases into two cohorts, as described above: the Migrated cohort contained 839 (42%) of the cases and the Resident cohort contained 1166 (58%) of the cases. This could introduce further selection bias since, as noted in section 2.2.2, migrants generally tend not to be representative of their region of origin.

Whether the Combined cohort of cases (Migrated plus Resident) should be considered in any particular analysis, or just the Migrated or Resident cohort, depends upon what is being analysed. If a measure is required for the burden of disease within New Zealand then the full cohort should be used. If, however, the aim of the analysis is either to study, or to control, the effects of external migration then the relevant Migrated or Resident cohort should be used. Since the aim of this study was to control the effects of external migration so that effects of latitude within New Zealand could be analysed, most of the analyses presented in Chapter 5 are based solely on the Resident cohort. Apart from analyses of gender and MS phenotype, any further assessment of the Migrated cohort to ascertain which other factors might affect the prevalence of MS within that cohort, and hence which factors might affect the full burden of disease within New Zealand, is left to a future study. In addition, it is not known what proportion of the general New Zealand population at any given time has remained resident within New Zealand since birth, as opposed to the proportion who have migrated at some point in their life. If the relative numbers differ from those in the Resident and Migrated cohorts analysed here then the residence criterion used to differentiate the two cohorts also highlights migration (or lack of) as a potential risk factor.

The reduced cohort numbers also increase uncertainties in the results, especially since the data will be further stratified as described in section 4.3. Issues with sample size will be discussed with the results in Chapter 5. In addition to the overall reduction in numbers, whereas the original cohort had 90 Māori cases, the final Resident cohort had just 6 Māori cases, and the Migrated cohort had 17 cases. This meant that stratification of the results by ethnicity would not be statistically viable.

Now that the cohorts of cases have been established, the periods of interest can be determined as pre-requisites for obtaining the census boundary and population data.

4.2.2 Life Period Geographies

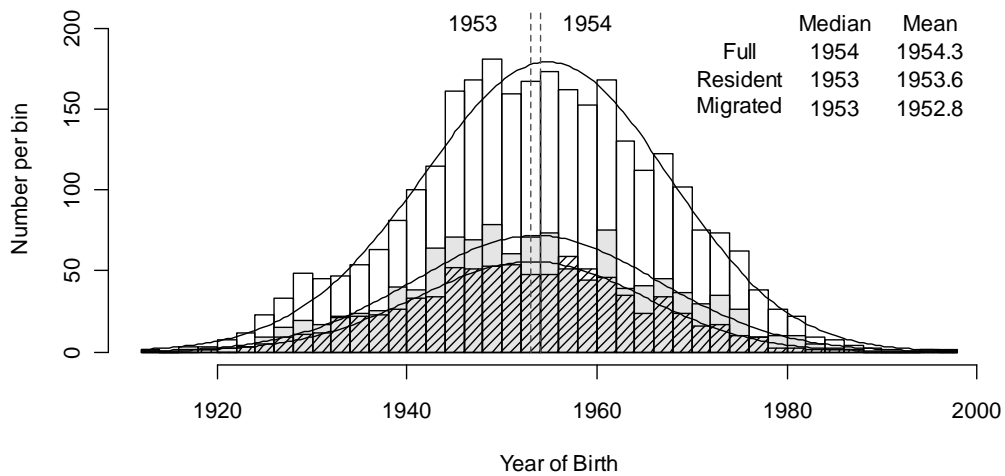
As described earlier in section 3.4, areal interpolation can be used to estimate the population distribution for a target geography as long as the distribution is known for some other source geography. The target geography used in this study will be discussed further in section 4.3; the most obvious source geography is that provided by the New Zealand population census. For prevalence on the 2006 census date, the

2006 census population data and data representing the census area boundaries are both freely available in digital format from Statistics New Zealand (StatsNZ n.d.), and thus provide a suitable source for areal interpolation. However, suitable distributions for other life periods of the MS cases being considered here cannot be obtained quite as easily.

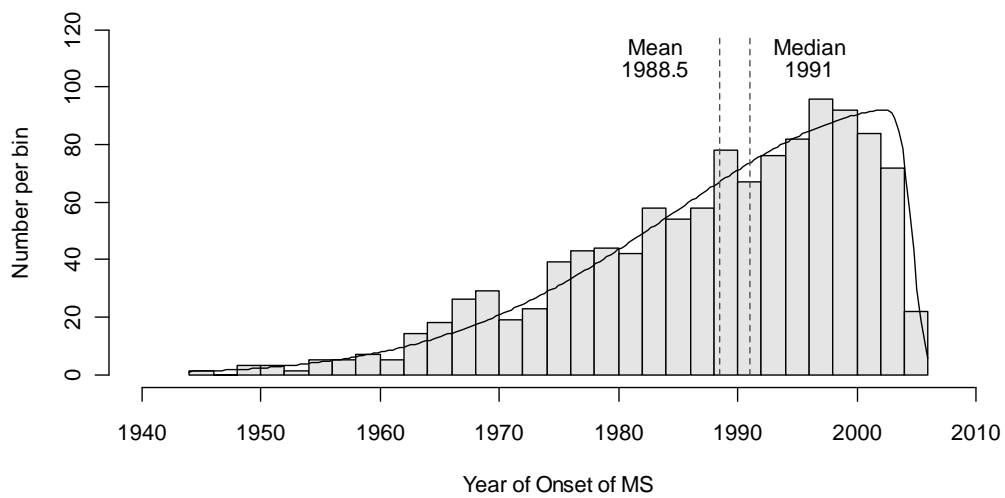
There are two main periods in the history of the MS cases in this study which are of significant interest: the period from birth and the period prior to onset of MS. The shaded distribution in Figure 4.1(a) shows the distribution of years of birth of the Resident cohort studied here, which has a mean of 1953.6. To align with census population age groups, the 0–4 age group (Age 0–4) is considered which has a mean of 1956.1. The census population data from 1956 is therefore taken as a representative population distribution for determining Age 0–4 prevalence. However, although the 1956 population data is available for download in digital format, the census boundaries are not. The methodology used for determining 1956 census boundaries is described in section 4.2.3. Also shown in Figure 4.1(a) are the distributions of years of birth for the full cohort of cases studied by Taylor et al (2010b) (unshaded) and the Migrated cohort of MS cases who had resided outside of New Zealand at some point prior to onset of MS (hatched). (Year of birth data was available independently of the questionnaire data.) It can be seen that the general shapes of the three distributions are similar and the choice of census population data would be the same for either the full cohort of cases, or the Migrated cohort. (All further discussion of cohort distributions in this section refer to the Resident cohort, unless stated otherwise.)

The other main period believed to be significant for MS (in common with most diseases) is the period prior to onset. The distribution of years of onset of MS is shown in Figure 4.1(b). There is a broad distribution, as with the distribution for years of birth, with a mean of 1988.5, though in this case the distribution is skewed to later years rather than being fairly symmetrical as with the years of birth. This would put the mean for the five years prior to onset of MS at 1986. Since 1986 is a census year in New Zealand, suitable census population data is available. However, as for the 1956 census, 1986 boundary data is not available in digital format.

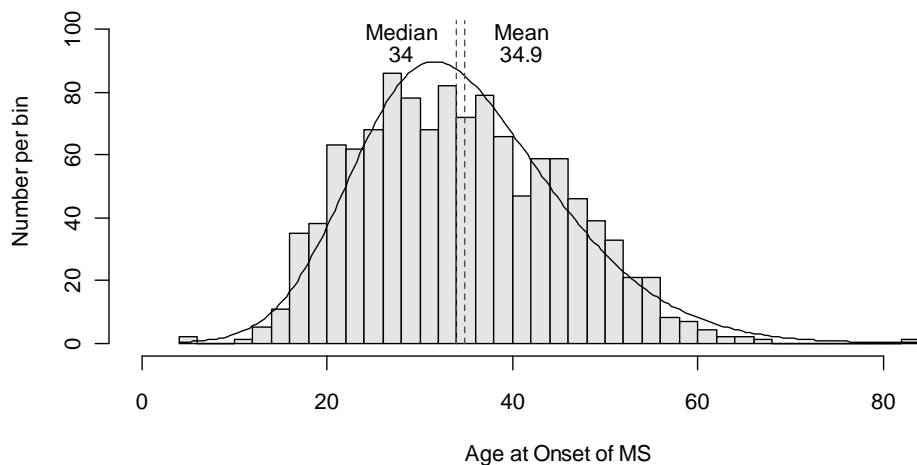
As described in the next section, 4.2.3, this study first looked at establishing a suitable population distribution based on the 1956 census data. Given the amount of time



(a)



(b)



(c)

Figure 4.1 Distributions of years of birth and onset, and age at onset of MS
 Distributions of: **(a)** years of birth; **(b)** years of onset of MS; **(c)** ages at onset of MS. Shaded: the Resident cohort, resident in New Zealand from birth to onset of MS. Unshaded: the full cohort of MS cases reported in Taylor et al (2010b). Hatched: the Migrated cohort, not resident in New Zealand at some point prior to onset of MS.

taken for this process and the time restrictions on this study as a whole, it was then decided not to repeat the process for the 1986 census data at the current time. Instead, results presented in Chapter 5 are restricted to prevalence for Age 0–4 and prevalence for the 2006 census day. Given that the distribution for years of onset of MS is skewed significantly to later years, it might be expected that prevalence results which would have been obtained for the period prior to onset of MS might be similar to those obtained for prevalence on census day 2006, as long as there had been no significant internal migration between onset and census day 2006. However, results presented in Chapter 5 suggest that this may not be the case.

The distribution of ages at onset of MS is shown in Figure 4.1(c). It can be seen that the age of onset of MS covers a broad range of ages from under 10 to over 80, with a slight skew towards lower ages. This confirms that MS is a disease of young people as well as older people, and suggests that triggers for the onset of MS are not strictly age related.

4.2.3 Census Boundaries and Population Data

More recent New Zealand National Census geographies use a spatial structure similar to that used in many other countries, including the UK: a hierarchy of Area Units (AUs) is built upon an underlying structure of meshblocks (MBs), as shown in Figure 4.2 for the 2006 census. Published census data is data aggregated into this hierarchy and, with suitable safeguards for confidentiality, made available for download from Statistics New Zealand (StatsNZ n.d.) at a variety of spatial resolutions. As indicated in Figure 4.2, the 2006 census data used in this study had a total of 1927 AU areas, though over 100 of these correspond to offshore and inland water. AUs are the first (lowest) aggregation of MBs; other options would have been 145 Territorial Authority (TA) areas, or 41,376 MB areas. (Note that 'AU' is used here for 'Area Unit' following the convention used by StatsNZ; 'CAU' for 'Census Area Unit' is also frequently used in New Zealand.) All of the New Zealand maps presented in this dissertation are based on data downloaded from StatsNZ.

Although the 1956 census population data was available for download in digital form from StatsNZ, a hard copy which included Volumes I to V of the eight published volumes was obtained through Inter Library Loan (NZDoS 1959). This gave summary information not available with the downloads and also small scale maps depicting the census areas used. Digital copies were made of the maps and relevant tables for cross

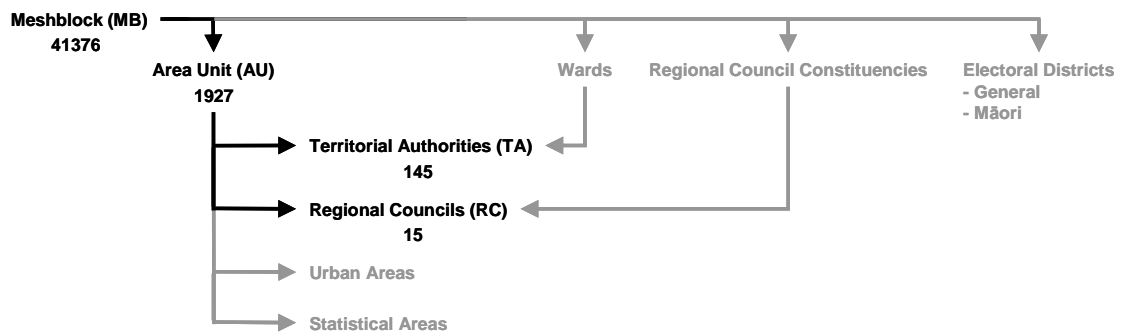


Figure 4.2 Geographic Hierarchy for the 2006 NZ Census

The geographic hierarchy for the 2006 New Zealand National Census indicating the boundary sets which are available within the standard meshblock dataset (black) and other census administrative areas (grey). Figures refer to the number of areas at that level of the hierarchy.

referencing with the electronic data. Two of the maps are shown in the next section and some of the tables are reproduced in Appendix B. The methodologies involved in combining the digital and hard copy data to provide 1956 equivalents to the digital 2006 data are described in sections 4.2.3.1 and 4.2.3.2. This process established boundaries for 305 areas at the city/town/county/borough level.

Ideally, the population data used in this study would be by gender, age group and ethnicity. However, the breakdown of the 2006 data by ethnicity was only into 10 year age groups, not the five year age groups which would be ideal. Additionally, the 1956 population data by area was only broken down into age groups or ethnicity, not both. A breakdown of the census population data by ethnicity was therefore not considered. This meant that all prevalence figures would be related to the population as a whole; using the non-Māori population in the calculations might be considered more suitable.

Although data for 2006 was readily available for download, boundary data for the 1956 census was not available at all, and the 1956 population data required some analysis before it could be used. Sections 4.2.3.1 and 4.2.3.2. now describe the techniques used to establish suitable data which could be used in further analysis.

4.2.3.1 The New Zealand 1956 Census Boundaries

Whereas the 2006 census areas were available through a digital download, the 1956 areas were only available from the 1956 census publication which included several maps with outlines of the County statistical areas. Rather than attempt to digitise all of these boundaries from the maps, the following fact was taken into account: although census boundaries are liable to change from one census to the next, many of the

boundaries tend to remain the same. Therefore, the clearest of the 1956 census maps were copied and used as a backdrop behind the 2006 meshblock areas. Separate maps of the North and South Islands of New Zealand are shown in Figure 4.3 and Figure 4.4. (Although these maps illustrate 'Movement of Rural population', this is not what is of interest here; here it is simply the boundary definitions which are required.)

After aligning the images with the outline of New Zealand within the GIS software, when compared with the meshblock boundaries, there were obvious distortions in the 1956 map image due to the original reproduction and subsequent scanning processes. However, after spatially adjusting (georeferencing) the map image, to allow for these distortions, the scanned County boundaries were found to fall very close to the boundaries of 2006 meshblocks in a very high proportion of cases. Given that many of the meshblocks are derived from traditional administrative boundaries, it seemed reasonable to attempt to build up the old 1956 census statistical areas from the current 2006 meshblocks. This avoided having to rely on directly digitising areas from either the inaccurate map from the census publication or from other electronic or hardcopy maps from the period. The complete list of census areas which needed boundaries was determined from the census population data as described in the next section, 4.2.3.2. The process used to construct the 1956 census boundaries is outlined in Appendix B.

Although some of the resulting boundaries would be deemed historically inaccurate, as a tool for placing the 1956 population prior to the areal interpolation process of disaggregation and then re-aggregation of the population distribution, this method was likely to introduce fewer inconsistencies than if the census areas were digitised manually from the census maps, or omitted from consideration completely. Indeed, despite limitations of the methods used, over 50% of the constructed 1956 areas had sizes within 0.2 km^2 of the published sizes and over 90% were within 5%. It was also noticeable that most of the high percentage discrepancies occurred with the smaller census areas since the smaller the area, the harder it was to match the size within a reasonable percentage. The areal interpolation processes will be described later in section 4.3. First it is necessary to consider the population data which corresponds to the census areas.

MOVEMENT OF RURAL POPULATION

NORTH ISLAND

The 1951-56 intercensal movement of population (including Maoris) in each county, exclusive of any city, borough, or town district which may be situated within its boundaries, is shown by the per cent increase or decrease on the 1951 figure, according to the following scale:—

	Increase	Decrease
Under 5%		
5-10%		
10-20%		
20-30%		
30% or over		

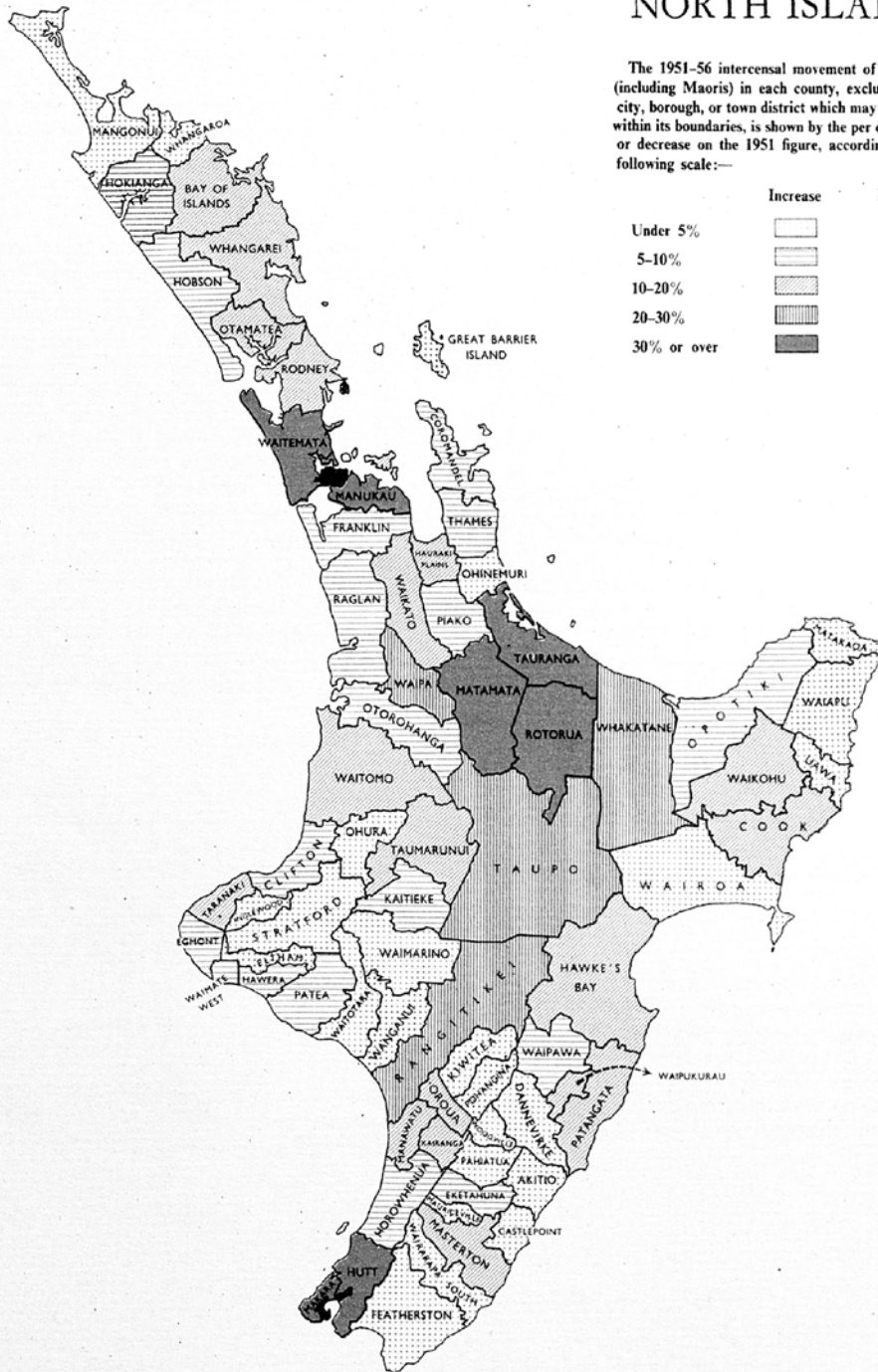


Figure 4.3 1956 census areas of North Island New Zealand

The North Island census areas from the 1956 New Zealand Census publication (NZDoS 1959) shown at approximately original size.

MOVEMENT OF RURAL POPULATION

SOUTH ISLAND

The 1951-56 intercensal movement of population (including Maoris) in each county, exclusive of any city, borough, or town district which may be situated within its boundaries, is shown by the per cent increase or decrease on the 1951 figure, according to the following scale:—

	Increase	Decrease
Under 5%	[White box]	[White box]
5-10%	[Horizontal lines]	[Horizontal lines]
10-20%	[Diagonal lines /]	[Diagonal lines \]
20-30%	[Vertical lines]	[Vertical lines]
30% or over	[Dark grey box]	[Black box]

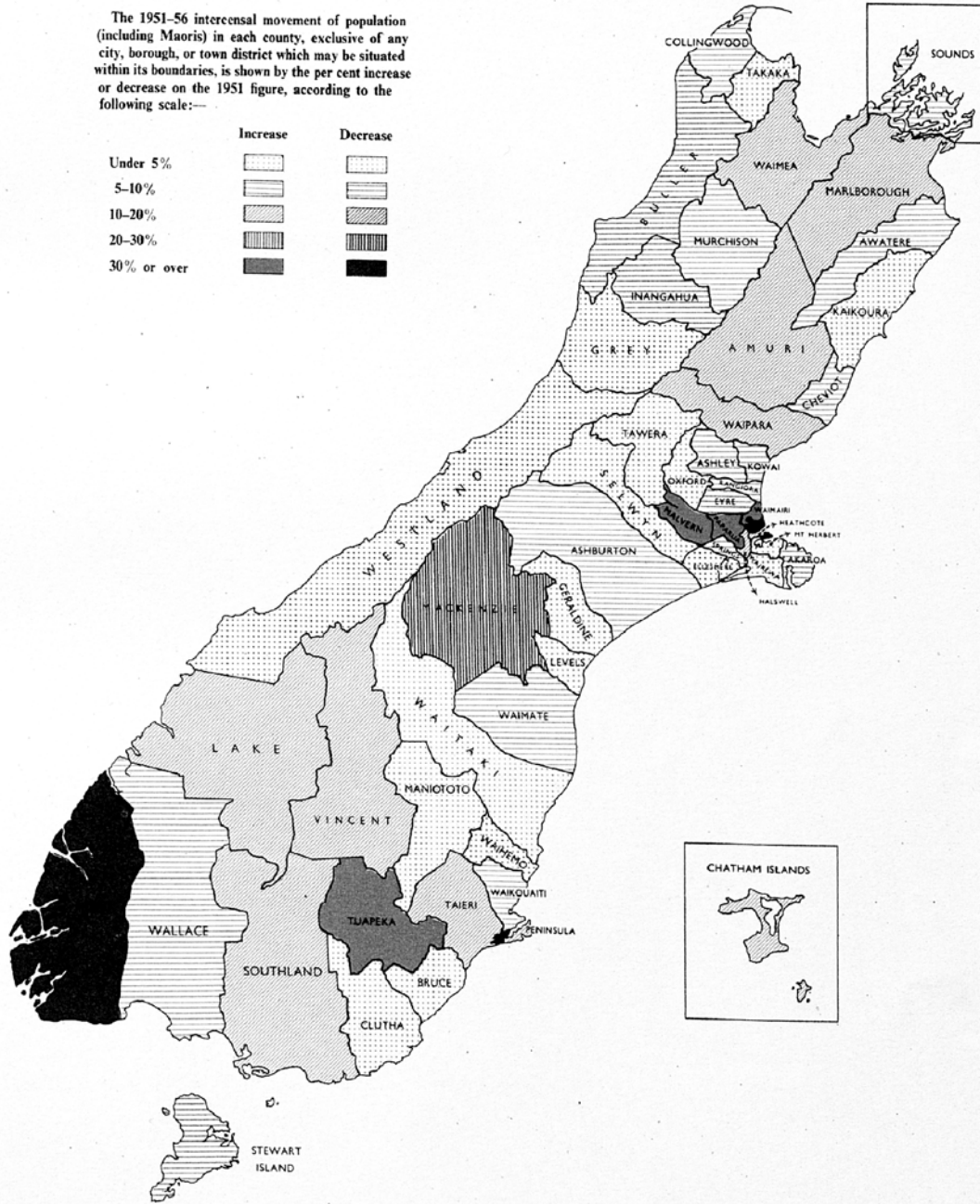


Figure 4.4 1956 census areas of South Island New Zealand

The South Island census areas from the 1956 New Zealand Census publication (NZDoS 1959) shown at approximately original size.

4.2.3.2 The New Zealand 1956 Census Population Data

As mentioned above, the census publication which included Volumes I to V of the eight published volumes was consulted (NZDoS 1959). In addition to the maps described in the previous section, various summary figures not present in the electronic downloads, indicated which census data should be combined to provide the full population data. In the census publication the data required was distributed across several tables, some giving a split by gender, others by gender and age. (Many tables also gave a split by Māori or non-Māori; such splits are not detailed here.) In all cases totals were also given which allowed for validation of data by cross-referencing with other tables. The main tables of concern here are given in Table 4.1, along with short table codes which will be used throughout the rest of this section.

Table VI_14 was useful as it indicated which County statistical areas the Boroughs and Town Districts were located within. The three tables listed from Volume II gave the population data required for this study. In addition, table VI_10 and the summary totals at the end of table VII_9 indicated how the population data for the various areas should be combined. Copies of these two tables plus table VII_11 are reproduced in Appendix B.

Volume I Table 10 (VI_10): [Split by gender]

Population of Administrative Counties (Including Dependent Town Districts)

Volume I Table 14 (VI_14): [Split by gender]

Population of Counties with their Interior Boroughs and Town Districts

Volume II Table 9 (VII_9): [Split by gender and age group]

Counties (Excluding Interior Boroughs and Town Districts)

Volume II Table 10 (VII_10): [Split by gender and age group]

Cities and Boroughs

Volume II Table 11 (VII_11): [Split by gender and age group]

(a) Town Districts not forming parts of Counties (Independent Town Districts)

(b) Town Districts forming parts of Counties (Dependent Town Districts)

Table 4.1 1956 Census Tables

The main tables from the Census publication used within this study. Short table codes are given in parentheses after the volume and table names.

	Administrative County	Boroughs and Cities	Independent Town Districts	Dependent Town Districts
<i>Hardcopy</i>				
VI_10	✓			✓
VI_14	✓	✓	✓	✓
VII_9	✓			
VII_10		✓		
VII_11(a)			✓	
VII_11(b)				✓
<i>StatsNZ</i>				
All TLAs	✓	✓	✓	

Table 4.2 Data available from the 1956 Census: Hardcopy and Electronic

Data available from various tables in the hardcopy census publication and electronic downloaded data. Codes for the census publication tables are as given in Table 4.1.

A summary of which data was included in which table is given in Table 4.2, along with the equivalent information for the electronic version of the census data from StatsNZ. The most important point to note is that the dataset for all Territorial Local Authorities (TLAs) from StatsNZ, only included the Administrative Counties, Boroughs, Cities and Independent Town Districts; the Dependent Town District data was not included, nor available for separate download. It would appear that this is due to a difference in data reporting in the original census publication which has not been carried over into the electronic version, as illustrated in the following example.

From table VI_14, Bay of Islands County consists of Kaikohe Borough, Kawakawa (Independent) Town District and Russell (Dependent) Town District, with individual and total populations as given in Table 4.3(a). The equivalent population data available from the 'All TLAs' data downloaded from StatsNZ is also given, though this data does not include Russell Town District. An indication of why Russell Town District has been omitted can be surmised from a comparison of tables VI_10 and VII_9, as shown in Table 4.3(b).

Both tables VI_10 and VII_9 report population by County, however, table VI_10 states that: "Town Districts forming parts of counties have been included with the counties in which they are situated" whereas table VII_9 lists "County (Exclusive of Interior Boroughs and Town Districts)". That is, table VI_10 includes Dependent Town Districts; table VII_9 does not.

	Males	Females	Total
<i>VI_14</i>			
Bay of Islands County	6,079	5,614	11,693
Kaikohe Borough	1,170	952	2,122
Kawakawa Town District (I)	358	400	758
Russell Town District (D)	308	309	617
TOTALS	7,915	7,275	15,190
<i>StatsNZ</i>			
Bay of Islands County	6,079	5,614	11,693
Kaikohe Borough	1,170	952	2,122
Kawakawa Town District	358	400	758

(a)

	Males	Females	Total
<i>VI_10</i>			
Bay of Islands County	6,387	5,923	12,310
<i>VII_9</i>			
Bay of Islands County	6,079	5,614	11,693
<i>Difference</i>	308	309	617
<i>VI_14</i>			
Russell Town District (D)	308	309	617

(b)

Table 4.3 Bay of Islands population data from the 1956 Census

Data for Bay of Islands County and Interior Borough and Town Districts.

(a) Table VI_14: Russell Town District is part of Bay of Islands County; StatsNZ: Russell Town District is not included. **(b)** Differences between tables VI_10 and VII_9 match Russell Town District in table VI_14; see text for details. Codes for the census publication tables are as given in Table 4.1.

Although example data is presented here for one County and its interior Borough and Town Districts, this omission of the Dependent Town District data is consistent across all 13 Counties which have Dependent Town Districts. The StatsNZ data would appear to be reporting data for Counties as if the data for the Dependent Town Districts was included whereas it is not included. Therefore, the additional data for the Dependent Town Districts was included from the census publication table VII_11. Without the hardcopy census publication tables, it would have been easy to have overlooked the separate Dependent Town District areas in the boundary definitions and to have only included the areas listed in the downloaded data.

Now that the source MS and census population data has been described, this chapter will continue with a discussion of techniques required for the analysis of that data.

4.3 Analysis Techniques

This section will outline technical aspects of the methodologies used in this study. The main software which was used was ArcGIS version 9.3.1 (ESRI 2009) which comes with the open source software python, version 2.5 (van Rossum and PythonLabs Team 2007). Although python was installed as part of the ArcGIS software, it can also be downloaded and used independently from ArcGIS, though version 2.5 is required if it is to be linked to ArcGIS 9.3.1. The open source statistics software R, version 2.10.1, was used for statistical analyses and graphical output (R Development Core Team 2009). R could either be used as a standalone program or it could be called from python using the python package rpy2 (Gautier 2009). In order to link all of these software programs through the python scripts it was also necessary to restrict the version of R used—changes in the way R 2.13 is installed, for example, means that it cannot be used from the version of rpy2 which works with python 2.5. If the methodology employed here is extended to ArcGIS version 10, which uses python 2.6, then a later version of R could probably be used but an alternative may be needed to replace rpy2. Information and links to the source software distributions for python, R and rpy2 can be found through the following web sites:

python: <http://www.python.org/>

R: <http://cran.r-project.org/>

rpy2: <http://rpy.sourceforge.net/>

In addition to the basic R software, various other add-on packages were used as part of this study. Reportlab was used for generating PDF files (Robinson et al. 2010) and the utilities xlutils, xlrd and xlwt were used for exporting data to Microsoft Excel format (Machin 2009a, 2009b; Withers 2009). The sn package (Azzalini 2010) was used for generating fits to various distributions and will be discussed further in section 4.3.4. Before that, section 4.3.1 will first outline the GIS techniques which were used to process the data. Sections 4.3.2 and 4.3.3 will continue by discussing the numerical techniques of age standardisation and least squares regression, which are used to determine prevalence rates of MS.

4.3.1 GIS Techniques for Aggregating and Processing Data

The requirement for areal interpolation of the source census population data was described in Chapter 3 and the choice of source data was discussed in section 4.2.3. Rather than use the population weighted centroids of the source data to represent the latitude component of the data, the target geography, or spatial reference, chosen for this study was a series of regular 'slices' at constant latitude, with either the central latitude or the population weighted centroid of the slice used to represent the slice. (See, for example, Figure 4.5(b), below.) This choice allows for adjustment of the resolution of the interpolated data according to the data available without resorting to a complete change in the methods employed for each resolution. These methods will be described in section 4.3.1.1. Since the process was required to be carried out by gender and age group, the ArcGIS ModelBuilder was initially used in order to provide an element of automation in the process. The ModelBuilder was then replaced by the direct use of python scripts from within ArcGIS. These tools are described further in Appendix C.

In addition to the pre-processing of the census population data, ArcGIS was also used to establish which of the MS cases were resident in New Zealand at given times in their lives so that the total sample of MS cases could be subdivided according to the residence requirement of the life period being considered. For reasons explained in section 4.2.1, two cohorts of cases were analysed, a Migrated cohort and a Resident cohort. The residence requirement used here for the Resident cohort was that cases were resident in New Zealand for their whole life from birth to onset of MS.

Further processing of the MS data to provide prevalence data was also carried out using python scripts from within ArcGIS. (See Appendix C for more details). This was for convenience—the scripts could take the relevant data directly from the ArcGIS geodatabases where the preprocessed data was stored, process it and store the processed data back into suitable database tables. Further python scripts could then take the processed data and either produce statistical and graphical output using the statistics software R directly from the python scripts, or export the data into a format suitable for using in Microsoft Excel or R outside of ArcGIS.

4.3.1.1 Using ArcGIS for Areal Interpolation

The process involved for the areal interpolation of the census population data and the census boundaries into the target latitude 'slices' was broadly as follows and is illustrated in Figure 4.5 for the 2006 Regional Council (RC) areas, or regions:

- Create a grid of latitude 'slices' of the required size and large enough to cover the extent of New Zealand being considered (166°E 48°S–180°E 34°S)
- Assume that the population density within each census area is uniform and assign the density to each area
- Use ArcGIS to 'slice' the census areas with the latitude grid
- Calculate the population of each sliced area by considering the area of the sliced area and the population density from the original census area
- Aggregate the sliced areas and populations into slices corresponding to the latitude slices of the grid being used

Note that Figure 4.5(c) suggests that the same latitude slices could be obtained simply by slicing the outline of New Zealand without consideration of the internal census areas. In terms of just the census area boundaries, this is correct. However, this would not take account of variations in the population density across the country which is used to apportion the population across the different latitude slices. Once the population of each slice was established, that population was assigned either to the mid-point latitude of each slice, or to its population weighted centroid.

[The next section continues after the following Figure.]

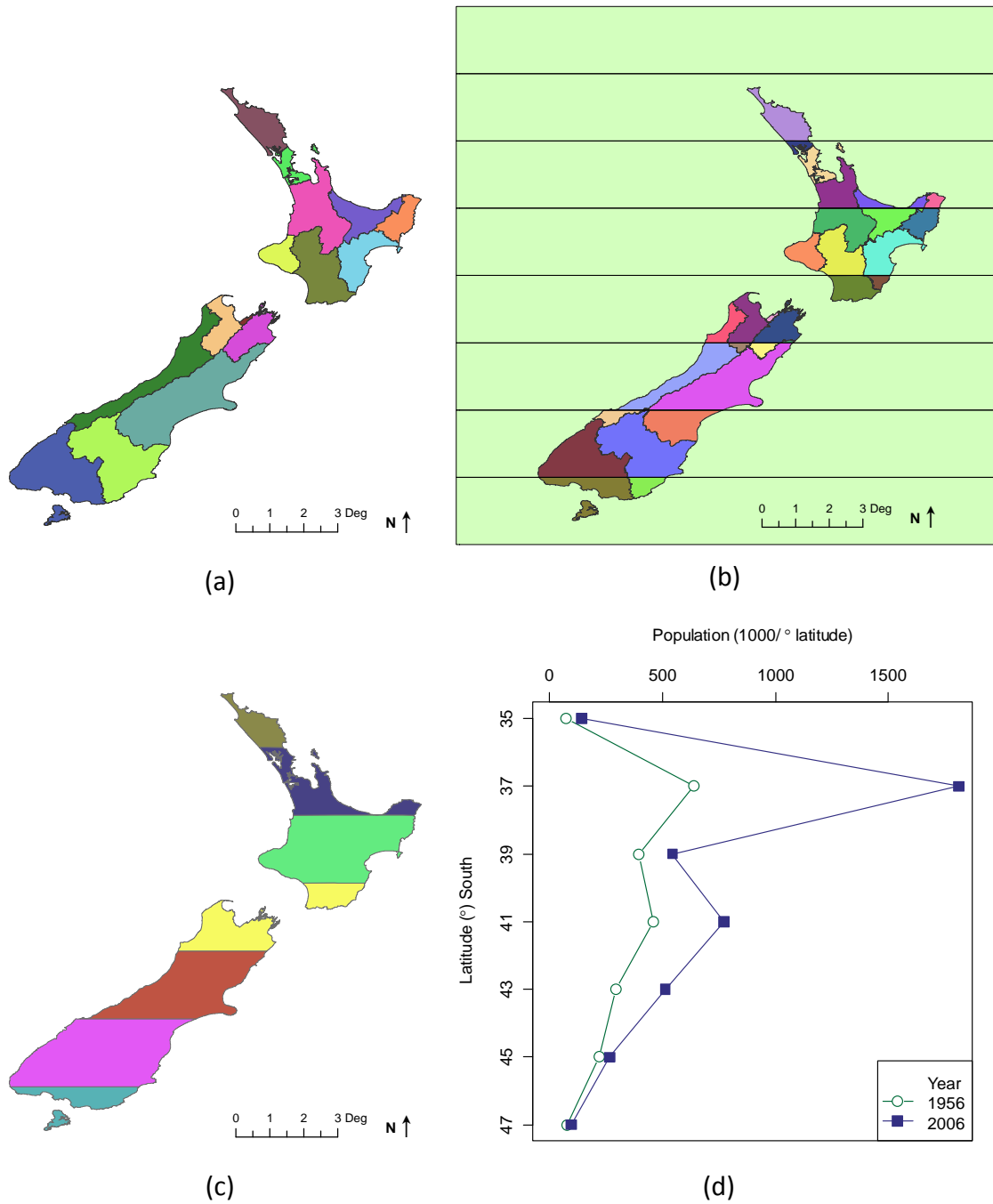


Figure 4.5 Areal interpolation of New Zealand 2006 census areas

The areal interpolation used to slice the New Zealand census areas into latitude slices. Shown here are 2° slices. For clarity, the areas are the 15 RC areas for the 2006 census rather than the 1799 AU areas; populations and population densities are not shown.

(a) The 15 regions within New Zealand. Each region is assigned its own population density. **(b)** The regions divided into sub-regions along the lines of the latitude grid, shown in the background. Each sub-region population is calculated from its area and the population density of the original region. **(c)** Sub-regions are combined into slices according to latitude. Each slice population is the sum of the sub-region populations of the slice. **(d)** 2006 population distribution, based on the 1799 AUs and the 1956 population distribution.

4.3.1.2 ArcGIS ModelBuilder and Python Scripts

Any data analyses which were to be made would need to be carried out for a number of different factors, including gender and MS phenotype. Therefore, it was decided to automate as many of the analyses as possible. This would also enable reproducibility of results and facilitate further factor analysis as suggested by initial results. The overall process and data workflow established for the analyses used in the investigation of prevalence with latitude is summarised in Figure 4.6. As well as allowing for factors such as MS phenotype, an ability to pre-filter the data was included so that several factors could be used in the stratification process. A method of pre-filtering was used rather than combining separate factors in order to focus the analysis on one particular factor.

ArcGIS processing can be scripted using the ArcGIS ModelBuilder and the scripting language Python. Considerations and methods used are outlined in Appendix C. Each of the processes in Figure 4.6 was automated with a Python script; processes could also be linked using further scripts. The use of scripts also facilitated consistency of naming and the systematic storing of data and results. The main python scripts used in this study are listed in Appendix C.

Having outlined GIS techniques, details of numerical techniques will now be given.

4.3.2 Age Standardisation of Data

The process of age standardisation aims to normalise rate data in such a way that the resulting rate represents that which would be obtained in a standard population—it aims to reduce bias in the results due to differences in the underlying population demography. Although this process utilises a *standard population*, there is not just one standard population and there has been much discussion as to which standard should be used (Ahmad et al. 2001; Robson et al. 2007; Zivadinov et al. 2003). This study has a sample which is predominantly of European ancestry so the European standard has been used. In addition, the Segi world standard was considered and variations in the results obtained are discussed in section 5.2.2. The European and Segi standards are described in more detail in section 4.3.2.1 and the method for calculating age-standardised rates is then given in the following section, 4.3.2.2.

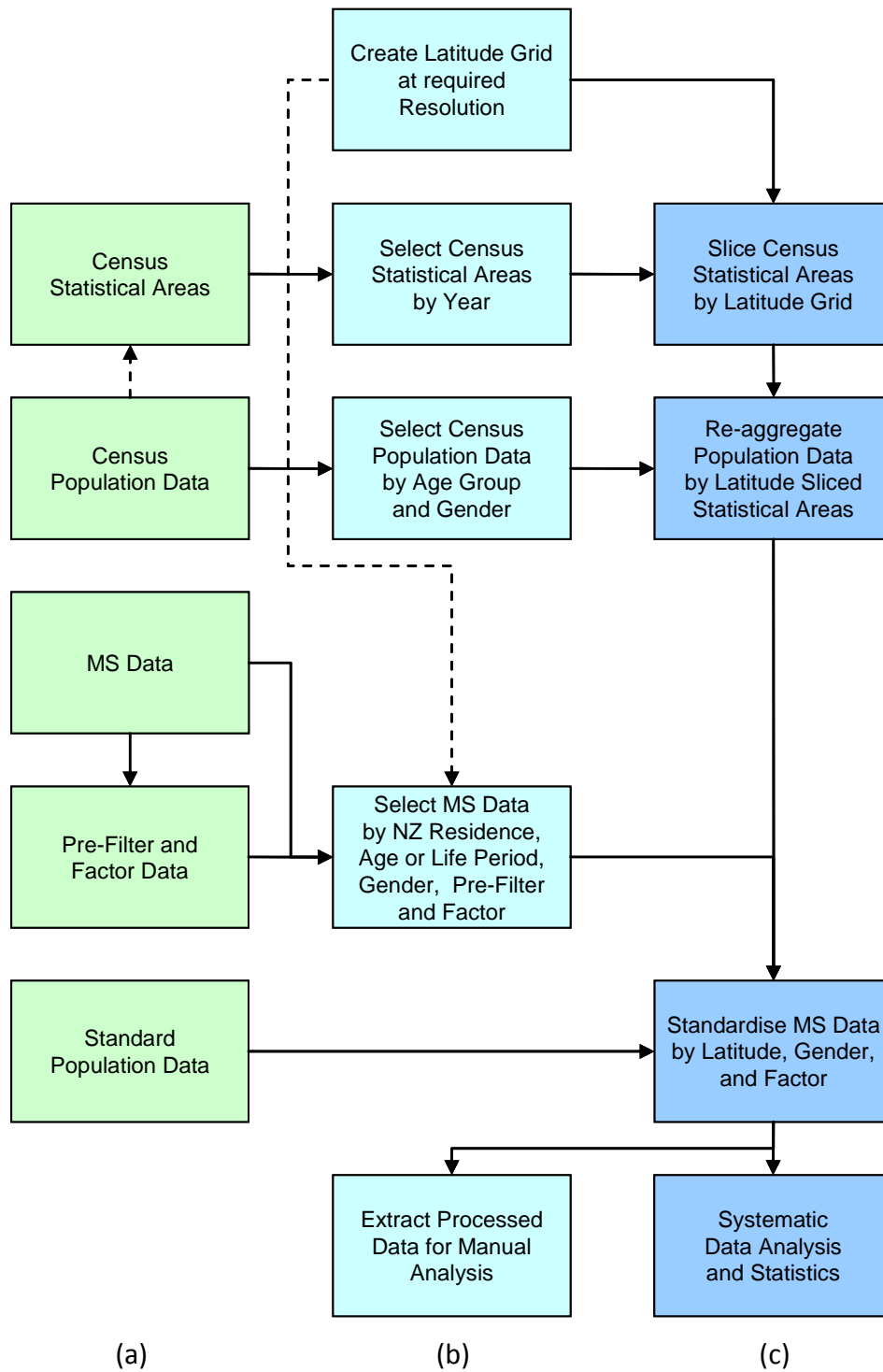


Figure 4.6 The process and data flow for the analyses

The process workflow can be split into three sections as shown. Arrows indicate the direction of the flow of data. **(a)** Census data and pre-processed MS data is manually reformatted into files or ArcGIS tables suitable for further processing. **(b)** Scripts select data according to the required criteria and output processed data into files or ArcGIS tables suitable for further processing. **(c)** Scripts combine and process data from various ArcGIS tables and systematically produce graphical and statistical output.

4.3.2.1 Standard Populations

A standard population is effectively just a representative age structure of some population. In most, if not all cases, the total population used is 100,000. This is often used for convenience: since many rates in epidemiology are given as a rate per 100,000 using this total avoids another factor in various equations.

Until recently, the most commonly used world standard population was the Segi standard population put forward by Segi (1960) and modified by Doll et al (1966). This was based on the age structures of populations taken from 46 different countries. The Segi standard has only recently been surpassed by the newer WHO world standard, put forward in 2000 (Ahmad et al. 2001), and designed to better represent world populations in the years 2000–2025. The standard adopted as a European standard was based on studies of Scandinavian populations (Doll and Cook 1967), and is generally used for studies based within Western Europe. Since the data being considered here is of predominantly white, European ancestry, it seems reasonable to use this European standard here. For comparison of results, the older Segi world standard is also used. This is in preference to the new WHO world standard since the sample populations being considered are mainly from the mid to late 1900's, so a standard established in the 1960's is more appropriate than one established for the 21st century. It is also interesting to note that if there had been a significant Māori sample in the study then the different structure of the Māori population would have warranted a different choice of standard, as discussed by Robson et al (2007).

The best way of describing a standard population is by illustration. Figure 4.7 gives a comparison of the Segi and European standards. The Segi world standard shows an almost uniform decline from a high 0–4 age group to a low 80–84 age group. The European standard, however, is significantly different. The 0–4 age group is just above average, then all age groups up to 55 are the same. The population only starts to fall with the 55–59 age group at which point it falls steadily, as with the Segi standard. The 85+ age group is the same as the 80–84 age group and represents the total of all ages 85 and over. Although not illustrated here, the new WHO world standard falls between the Segi and the European standards. This is an acknowledgement that the global birth rate is falling and that people are living longer. The standard also extends the age groups out to 100+ rather than just 85+.

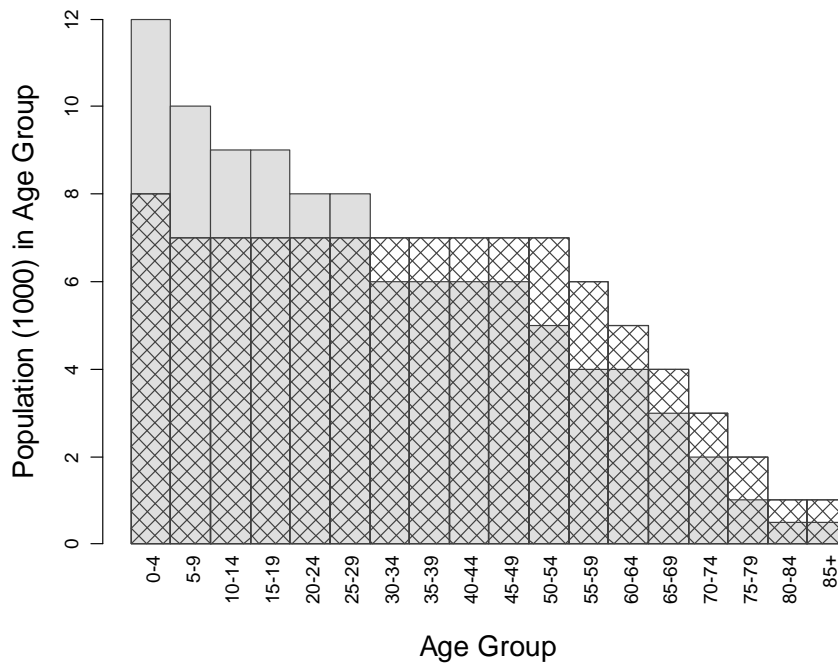


Figure 4.7 European and Segi World standard populations

The age structures of the Segi World (solid shaded) and European (cross hatched) standard populations (Ahmad et al. 2001). The new WHO world standard (not shown) lies between the two.

Now that the choice of standard population for the age standardisation process has been discussed, it is necessary to outline the process itself in order to establish how this choice of standard is combined with the census population data and the MS case data; the age standardisation process is described in the next section.

4.3.2.2 Calculating Age-Standardised Rates

The *age-standardised* rate is a weighted average of *age-specific* rates. For N age groups, the age-specific rate for each group i is d_i/y_i giving

$$\text{Age-standardised rate} = \sum_i^N \frac{d_i w_i}{y_i} \tag{4.1}$$

where d_i is the number of cases in the i th age group and y_i is the population size in the i th age group. w_i is the weight applied to the i th age group, being the population size in the i th age group of the standard population. It is also usual to have

$$\sum_i^N w_i = 100,000 \tag{4.2}$$

so that the age-standardised rate is given as the number per 100,000. By comparison, the crude rate, per 100,000, is given by

$$\text{Crude rate} = \frac{100,000 \sum_i^N d_i}{\sum_i^N y_i}$$

4.3

For this study, areas are defined by latitude and population subgroups are defined by gender and age, with other factors to be introduced at later stages. As such, the population used for the variable y_i is the census population within the given latitude area for the given gender and an appropriate year. The census year chosen is that which corresponds to the mean year of the sample population. For the prevalence on census day in 2006 this mean is, by definition, 2006. For the residence locations corresponding to Age 0–4, the mean is around 1956 so the 1956 census data is used, as discussed earlier in section 4.2.2. Note that if subsequent 5-year age groups, or the period prior to onset, were to be studied then this would necessarily require the census data from subsequent census years. As also discussed in section 4.2.2, further census years would not be considered in this current study.

The next section describes the least squares regression techniques which are used to determine prevalence–latitude gradients from the age-standardised data.

4.3.3 Regression Techniques

The variation in prevalence with latitude is often referred to as a latitude gradient, and the very nature of the data leads to it being modelled (fitted) using linear least squares regression. However, linear regression is not ideal for rate data: by definition the dependent variable (the rate) can only vary from 0 to 1 whereas for a linear fit it can exceed this range even if the independent variable is restricted. Although linear regression results are reported here, an alternative log regression is also used. The more immediate restriction of requiring the rate to be positive is enforced with log regression. Though, in theory the rate could still rise above 1, in practice, for the range of rates being considered here, this is not a problem. A second alternative, logistic regression, would restrict the full range of the dependent variable to [0,1] as required. However, logistic regression requires control data which is absent in this study.

Interestingly, for rates as low as those in this study, regression coefficients obtained for log and logistic regression would be almost identical, as shown in section D.2.

Note that any regression model assumes that the data is heteroscedastic, that is, that the variances in the dependent variable are similar for all values of the independent variable. In practice, however, higher values of the dependent variable will often have higher variance, at least in a linear model. In this situation, a log regression model will tend to even out the variances which is another reason for considering a log regression model over a linear model.

As mentioned at the start of section 4.3, statistical output was obtained using the statistics package R. As with many other statistics packages, R can report regression fit coefficients and carry out various tests to compare fits without requiring the user to have any knowledge of the underlying process. However, an understanding of the process can greatly help when it comes to interpreting the output and results obtained. In particular it is necessary to understand what is being tested and the concept of the null hypothesis.

The null hypothesis (H_0) is simply a statement of what condition is being tested to indicate no effect. In other words, if the null hypothesis condition is likely to be true by the test being used, then it is likely that the variable being tested has no effect in the model. The corollary is that if the null hypothesis condition is unlikely then the variable being tested is likely to have an effect in the model. The null hypothesis condition is usually written as $H_0 : \beta = 0$ (for example) with β being the variable under test, and $\beta = 0$ (in this example) being the test under consideration. The test is usually made using ANOVA tables to give the p -value.

Appendix D outlines various regression techniques and ANOVA in the context of this study. Rate data suggests the use of logistic (logit) regression, however, the lack of control data means that log regression must be used in preference. It is shown in Appendix D that either log or logistic regression are likely to give the same results within the uncertainties of the data and the other methods involved in this study, so log regression will be used in this study. For a log regression fit to prevalence–latitude data, the gradient β_1 gives a so called *rate ratio*, $\exp(\beta_1)$, which gives the relative change in prevalence per degree of latitude south. A brief discussion is also given in Appendix D on comparing two regression fits. For this, the null hypothesis test being

considered is whether the difference between the two gradients is likely to be zero. Results given in Chapter 5 include the p -value from such ANOVA F tests for 95% confidence intervals (CI) without further qualification. It is also shown in Appendix D that, for log regression, the F:M sex ratio must vary with latitude unless the rate ratios (or gradients) for the female and male data are the same. Therefore, whether the F:M ratio will change with latitude or not can also be implied from the same ANOVA test. The next section continues the discussion on regression techniques by considering the effects of the sample size and the size of the underlying census population.

4.3.3.1 Sample and Population Size Effects in Regression

In order to compare the trends obtained from two different samples it is important to take account of the sample size and how the data has been normalised (or age-standardised) to give the rate. As an example, consider the same sample of MS cases and compare the crude rates for 2006 and Age 0–4 (1956). Within the study sample, there were 1166 MS cases for whom full data was available and who lived in New Zealand in 1956 and remained within New Zealand at least until the onset of MS. Note that there were other MS cases who lived in New Zealand in 1956, but those cases are not included in this sample since they had either died or migrated out of New Zealand at some point prior to onset of MS, or they were not present in New Zealand on the 2006 census day. The total New Zealand populations for 1956 and 2006 were 2.16M and 4.14M, respectively, so the two crude rates for MS cases are 54 and 28 per 100,000 respectively. This apparent halving of the rates from Age 0–4 to 2006 is simply due to having a fixed sample and a population which almost doubled from 1956 to 2006. If the latitude gradients from linear least squares fits were compared for these two situations then there would also appear to be the same change in gradient, even if all of the cases remained in the same locations and there was in fact no change in the distribution of cases. (Effects of migration are discussed further in section 4.3.5). However, log regression gives the rate ratio which, for these two scenarios, would not change despite the factor of 1.9 in the linear gradients. In this case, the population effects are absorbed by the intercept which is eliminated when taking the ratio. (See D.2 for details.) To investigate the effect of latitude on the prevalence rates independently of the sample size and other total population effects, the log regression must be used and only the gradients (rate ratios) compared.

Although least squares fits can be calculated manually using the equations given in Appendix D, they can also be calculated automatically using statistics software such as R. The same software can also be used to generate graphical output as described next.

4.3.4 Statistics and Curve Fitting Using R

The software program R (R Development Core Team 2009) is a command line, scripting program based on the data analysis and graphics language S (Becker and Chambers 1984; Becker et al. 1988; Chambers 1998; Chambers and Hastie 1992), but provided under a 'GNU General Public License'. There is an extensive core package which includes statistics and graphical output routines, and greater functionality can be added through purpose built packages. Linear and log regression techniques are available as standard within R. A description of these and how to use R is beyond the scope of this dissertation and many results, such as those from ANOVA tables, will be presented without further explanation. However, as part of the analyses and presentation of results, various fits to data are used which are not covered in sections 4.3.3 and Appendix D on linear and log regression. A brief outline will be given here as background to fits presented throughout this dissertation.

Many natural processes are driven by random events and, as such, data often has an inherent uncertainty. This is usually seen in the form of a normal distribution—take the same measurement many times and the values obtained will often form a broad, symmetrical distribution such as those for birth years given in Figure 4.1(a). But many distributions are also skewed, as for the years of onset of MS shown in Figure 4.1(b) and the age at onset of MS shown in Figure 4.1(c). Routines for fitting normal and certain types of skewed distributions to data are intrinsic to R. However, there is another form of skewed distribution which has received much attention in the literature and which is also available as an add-on R package. Although similar distributions had been developed independently by a number of researchers, the distribution has been widely termed the "Azzalini skew normal" (SN) distribution (Arnold and Beaver 2002) due to the important contribution in this field of work by Azzalini (Azzalini 1985; Azzalini and Dalla Valle 1996). The SN distribution is available for analyses through the R package 'sn' (Azzalini 2010). A useful property of the SN family of distributions is that it naturally includes the normal distribution. A fit using the SN does not, therefore, preclude a normal distribution as the result. In addition,

the SN has been shown to be applicable to 'truncated' data in the same way that the normal distribution is applicable to unrestricted data (Arnold and Beaver 2000, 2002; Arnold et al. 1993). 'Truncated' here refers to some limit on the measuring or selection process. One example given in Arnold and Beaver (2002) is that of the distribution of waist sizes for uniforms of elite troops. The troops must satisfy a minimum height requirement which results in a positively skewed distribution of waist sizes. Here, in Figure 4.1(b) and Figure 4.1(c), the distributions are skewed and the SN fits provided as guides to the eye are reasonable fits. Although various causes can be argued for the skewed nature of these distributions, the use of the SN here does not imply any truncation of the data *a priori*. Rather, the use of this functional form is merely for convenience.

The final technique presented here is a simple method for establishing the general behaviour of a population (and sub-population) which is subject to internal migration.

4.3.5 Inter-Regional Migration

It is not the aim of this study to investigate or model population migration in any depth. However, as discussed in Chapter 3, migration does appear to have a role in the prevalence of MS. So if a model can be established which mimics population migration then that model could be applied to a sub-population to see how that sub-population might be re-distributed over time, assuming that it is subject to the same migration forces as the population as a whole. In the context of this study, if a person's susceptibility to MS is determined solely on residence location during pre-birth and early years, then the distribution of cases of MS at later years will be determined by the early prevalence distribution, confounded by migration effects. The aim of this section, and subsequent results reported in Chapter 5, is to investigate the confounding effect of internal migration and to compare the modelled effect with any effects seen in the cohort of MS cases studied here.

It was mentioned in Chapter 3 that matrix methods can be used to study migration. In particular, inter-regional migration can be modelled easily in this way. Appendix E puts forward a simple matrix analysis in which it is assumed that there are no births or deaths and that there is no migration into or out of the regions from or to areas outside. This may seem somewhat restrictive, however, it does correspond to the residence criteria for the Resident cohort of MS cases being looked at in this study,

though the population as a whole would not be restricted in this way. Given a starting population structure for a number of nominal regions, the model gives a predicted population structure after a specified number of time steps. (Note that each time step in this model is arbitrary and not necessarily equivalent to a year.) In addition, it provides a migration matrix which can be applied to any sub-population to determine how that structure would change given the same migration conditions. An initial population structure similar to that from the 1956 New Zealand census was chosen along with model parameters which gave a final structure similar to the 2006 census structure; these are shown in Figure 4.8(a). The migration matrix from this model was then applied to an initial MS prevalence–latitude profile similar to that seen in New Zealand. This initial profile and the final model output are shown in Figure 4.8(b).

It can be seen that there is a slight shift in the trend of the data after the migration matrix is applied—rates for the low regions, around the high peak in population, have increased and rates for the higher regions, which have generally lower populations, have reduced. Assuming that the regions here are 2° latitude apart and increasing region number corresponds to increasing (absolute) latitude, then log regression fits, as described in Appendix D give rate ratios for the initial and final profiles of 1.103 and 1.086 per degree latitude, respectively, a change of about 1.5%. That is, the prevalence

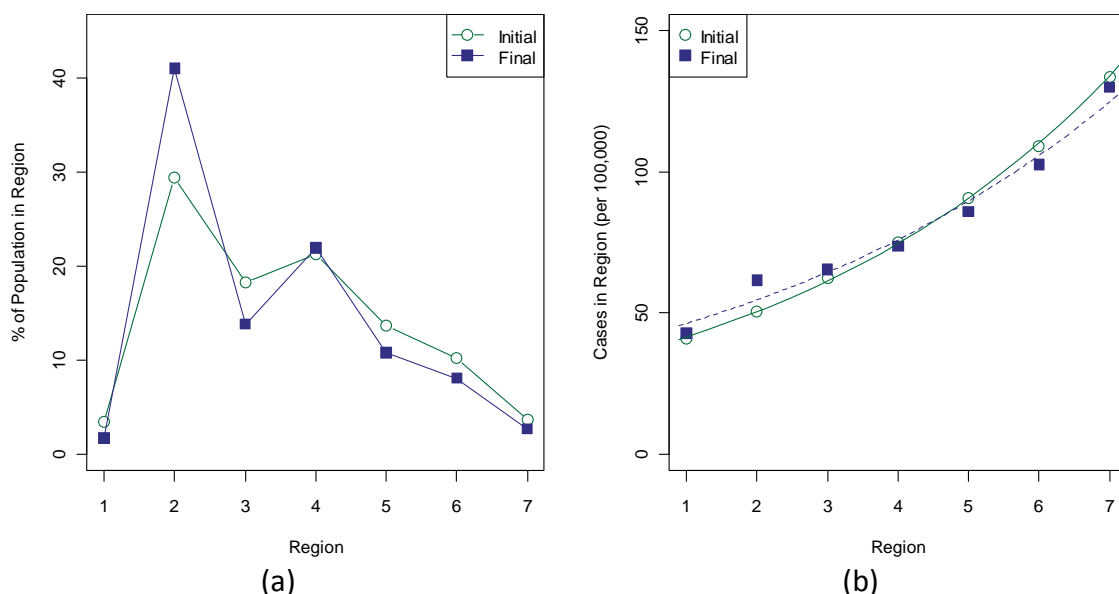


Figure 4.8 Modelling inter-regional migration

The population distribution and prevalence rates obtained after seven time steps of the matrix migration model described in the text. **(a)** Population distributions: initial based on the 1956 census population; final obtained from the model. **(b)** Initial prevalence rates and final prevalence rates after applying the migration matrix to the sub-population. The curves shown in **(b)** are log regression fits to the data.

of MS increases slightly faster with latitude for the initial population distribution than it does for the population distribution after model migration effects have been applied.

Although the effect observed in these model results is not a strong effect, it does highlight the following: if a similar shift in rate ratios is observed in the data being studied here then it will be necessary to consider how much of any effect is due to migration and how much due to non-migration effects. (This effect will also be confounded by birth, death and external migration factors not considered here.) If, however, a trend is seen which suggests a stronger effect (greater rate ratio) in the later data compared with the early years data, or a shift in the opposite direction, then it is likely that the effect is due to non-migration effects.

4.4 Summary of Methods

In order to study the effect of latitude within New Zealand on the prevalence of MS at different periods in the lives of the cohort of MS cases within this study, the cohort must be restricted to those cases who have been resident in New Zealand from birth to onset of MS (the Resident cohort). The distribution of birth years present within the Resident cohort requires Census data from 1956 in order for the distribution of MS cases from Age 0–4 to be studied. Although the data and GIS techniques are available to allow this analysis, setting up the data is a time consuming process. Therefore, the equivalent process for the years leading up to onset of MS was not carried out at this time and comparative analyses are only available for prevalence on census day 2006. In order to analyse the effect which external migration might have on the prevalence–latitude gradient of MS within New Zealand, those not included in the Resident cohort were included in a Migrated cohort. Analyses of the Migrated cohort would, however, be restricted to census day 2006.

Techniques are available which can automate the processing and analysis of the MS data using various GIS and analysis software programs. The choice of the European standard population for age standardisation of the data has been established, along with the degree of applicability of linear, log and logistic regression to the analyses required in this study, and the preference for log regression to provide rate ratios rather than standard linear regression gradients. A simple model for inter-regional migration can also be used to indicate how migration might confound the distribution of prevalence.

Chapter 5

Prevalence of MS with Latitude

5.1 Introduction

As stated in Section 1.2, the aim of this study is to establish the effect of latitude on the prevalence of MS in New Zealand and to examine whether prevalence rates vary for different periods in the life of the MS cases. This chapter presents the results from this study based on the analyses described in Chapter 4.

Section 5.2 gives an overview and summary results of the prevalence of MS with latitude within New Zealand for census day 2006 and for location of residence at Age 0–4. This summary is followed by a discussion of how the results are effected by the choice of areal interpolation technique and the choice of standard in the age standardisation process. Variations in prevalence with various factors such as gender and MS phenotype are then discussed in Section 5.3, along with a look at the effects of migration, followed by an examination of the effect of month of birth on prevalence. Section 5.4 concludes this chapter with a summary of the results presented.

Note that this chapter contains many figures and tables. Where possible, these are placed in context, however, in some places they are grouped at the end of relevant text in order to aid continuity of the text. Table 5.1 and Table 5.2 contain lists of tables and figures with page numbers which may aid the reader.

Table	Page	
Table 5.1	List of Tables in Chapter 5	123
Table 5.2	List of Figures in Chapter 5	124
Table 5.3	Prevalence of MS with latitude in New Zealand	125
Table 5.4	Gradients of prevalence of MS with latitude in New Zealand	128
Table 5.5	Linear and log regression gradients—mean latitudes	140
Table 5.6	Linear and log regression gradients—weighted latitudes	141
Table 5.7	Linear and log regression gradients—Segi Standard	145
Table 5.8	Numbers of cases of MS by Gender and MS Phenotype	146
Table 5.9	Linear and log regression gradients—Gender	150
Table 5.10	Linear and log regression gradients—MS Phenotype (Female)	157
Table 5.11	Linear and log regression gradients—MS Phenotype (Male)	158
Table 5.12	Rate Ratio Comparisons for PPMS Phenotype and 3° latitude slices	163
Table 5.13	Log regression gradients—2006 Prevalence for Migrated cohort	166
Table 5.14	Net Migration—Gender	168
Table 5.15	Net Migration—Gender and MS Phenotype	169

Table 5.1 List of Tables in Chapter 5

Figure	Page	
Figure 5.1	Variation in prevalence of MS in New Zealand—census day 2006	127
Figure 5.2	Variation in prevalence of MS in New Zealand—Age 0–4	128
Figure 5.3	MS case numbers and population—mean latitudes	130
Figure 5.4	MS case numbers and population—population weighted latitudes	131
Figure 5.5	Linear regression fits—mean latitudes for 2006 Prevalence	134
Figure 5.6	Log regression fits—mean latitudes for 2006 Prevalence	135
Figure 5.7	Linear regression fits—weighted latitudes for 2006 Prevalence	136
Figure 5.8	Log regression fits—weighted latitudes for 2006 Prevalence	137
Figure 5.9	Linear regression fits—weighted latitudes for Age 0–4 Prevalence	138
Figure 5.10	Log regression fits—weighted latitudes for Age 0–4 Prevalence	139
Figure 5.11	Rate Ratios—Latitude Slice Size	142
Figure 5.12	Variation of age structure with Latitude	144
Figure 5.13	Log regression fits—2006 Prevalence by Gender	148
Figure 5.14	Log regression fits—Age 0–4 Prevalence by Gender	149
Figure 5.15	Rate Ratios—Gender and Life Period	151
Figure 5.16	Log regression fits—2006 Prevalence by MS Phenotype (Female)	153
Figure 5.17	Log regression fits—Age 0–4 Prevalence by MS Phenotype (Female)	154
Figure 5.18	Log regression fits—2006 Prevalence by MS Phenotype (Male)	155
Figure 5.19	Log regression fits—Age 0–4 Prevalence by MS Phenotype (Male)	156
Figure 5.20	Rate Ratios—Life Period, Gender and MS Phenotype	159
Figure 5.21	Rate Ratios—Gender and MS Phenotype (2006)	160
Figure 5.22	Rate Ratios—Gender and MS Phenotype (Age 0–4)	161
Figure 5.23	Log regression fits—2006 Prevalence for Migrated cohort	165
Figure 5.24	Net migration within New Zealand—Pre Onset	171
Figure 5.25	Net migration within New Zealand—Post Onset	172
Figure 5.26	Number of births in New Zealand by month of birth	174
Figure 5.27	Crude prevalence by Month of Birth	175

Table 5.2 List of Figures in Chapter 5

5.2 Overview of Prevalence of MS in New Zealand

The prevalence rates of MS in New Zealand for census day 7 March 2006 vary with the latitude of residence of the MS cases. Summary results of crude prevalence and prevalence age-standardised to the European Standard are presented in Table 5.3(a) for 2005 of the 2917 MS cases in the NZMSPS project (the 2005 cases for which the actual residence location on census day 2006 was disclosed to the survey). The age-standardised prevalence rate of MS was 31.4 per 100,000 in the far north of New Zealand and 100.1 per 100,000 in the far south. As can be seen in Table 5.3, the crude prevalence rates are within standard errors of the age-standardised rates. These prevalence rates are for 3° latitude slices, with each slice represented by its population weighted centroid, and are illustrated in Figure 5.1(a). A log fit to the data, shown in Figure 5.1(b), gives a rate ratio exponent, β_1 , of 0.111 ($\sigma=0.006$) with $R^2=0.99$. The

Pop Wt Lat (°S)	Prevalence per 100,000			
	Crude	Std Err	Age Std	Std Err
35.67	35.39	4.49	31.37	4.01
37.36	33.96	1.26	35.36	1.32
40.81	54.45	2.36	54.34	2.36
43.77	73.58	3.43	69.98	3.28
46.06	98.39	6.53	100.10	6.67

- (a) Period: census day 2006; Sample size: 2005;
Cohort: Combined: NZ residence location known for census day 2006

Pop Wt Lat (°S)	Prevalence per 100,000			
	Crude	Std Err	Age Std	Std Err
35.67	15.98	3.02	13.78	2.62
37.36	16.11	0.87	16.85	0.91
40.81	22.41	1.51	22.29	1.51
43.77	27.41	2.10	25.99	2.00
46.06	33.37	3.80	33.47	3.83

- (b) Period: census day 2006; Sample size: 839;
Cohort: Migrated: residence location unknown, or not NZ for any time prior to onset of MS

Pop Wt Lat (°S)	Prevalence per 100,000			
	Crude	Std Err	Age Std	Std Err
35.67	19.41	3.33	17.59	3.04
37.36	17.85	0.91	18.51	0.95
40.81	32.04	1.81	32.05	1.82
43.77	46.17	2.72	43.99	2.60
46.06	65.02	5.31	66.63	5.46

- (c) Period: census day 2006; Sample size: 1166;
Cohort: Resident: NZ residence location known from birth to onset of MS

Pop Wt Lat (°S)	Prevalence per 100,000			
	Crude	Std Err	Age Std	Std Err
35.65	33.88	6.08	18.51	3.32
37.63	38.34	2.07	24.93	1.35
40.82	55.99	3.04	39.22	2.13
43.81	68.20	4.33	50.01	3.18
46.02	97.72	6.82	68.28	4.77

- (d) Period: Age 0–4; Sample size: 1166;
Cohort: Resident: NZ residence location known from birth to onset of MS

Table 5.3 Prevalence of MS with latitude in New Zealand

Prevalence of MS in New Zealand for census day 2006 and Age 0–4 for the NZMSPS project cohorts as indicated. Data is aggregated to 3° latitude slices, given by the population weighted centroid (Pop Wt Lat) of each slice. Crude: crude prevalence; Age Std: prevalence age-standardised to the European standard; Std Err: standard errors.

rate ratio, $\exp(\beta_1)$, for this is 1.117, so, for the cases resident in New Zealand on census day 2006, the prevalence rate of MS increases by just under 12% for each degree of latitude south. The correlation coefficient, R^2 , for the log fit is slightly higher than that for the linear fit, which, combined with a simple visual inspection of the graphs, indicates that, for this data, the log fit appears to be more suitable than the linear fit. The regression fits to the data are summarised in Table 5.4.

Taylor et al (2010b) reported a prevalence–latitude gradient for the whole population of 10.7 ± 0.9 per 100,000 per degree of latitude south of 37°S. Scaled to the same sample size as here, this gradient would be 7.4 ± 0.6 per 100,000 per degree of latitude south of 37°S, and if the northern region was also included, the gradient would be around 6.1 per 100,000 per degree of latitude, which is slightly less than, but comparable to that given in Table 5.4, given the errors involved.

Table 5.3 also includes summary results for census day 2006 for the cohort of cases excluded from this study based on residency criteria (b), and cases resident in New Zealand from birth to onset of MS (c). Linear and log regression fits are illustrated in Figure 5.1(c)–(f) with rates and regression fit summaries given in Table 5.4. Crude prevalence rates cannot easily be directly compared due to different sample sizes, however, the rate ratios from the log regression fits can be compared.

For prevalence of MS on census day 2006, there is a significant difference in the rate ratios for the 'Resident' cohort of cases studied here, who were resident in New Zealand from birth to onset of MS, and the 'Migrated' cohort excluded from most analyses in this study, where residence location was either unknown or not in New Zealand for some period between birth and onset of MS. The rate ratios are 1.140 and 1.084, respectively, with a comparison of underlying gradients giving an Anova p -value of 0.004. So, for the Resident cohort of cases, resident in New Zealand from birth to onset of MS, the prevalence rate of MS increases by around 14% for each degree of latitude south. In contrast, for the Migrated cohort, which experienced migration to or from New Zealand at some point prior to the onset of MS, the prevalence rate of MS increases by only 8% for each degree of latitude south. (The 1.117 rate ratio for the combined cohort is seen to be a weighted average of the 1.140 and 1.084 rate ratios for the two constituent cohorts.) External migration is therefore a significant factor in establishing the effect of latitude on the prevalence of MS in New Zealand.

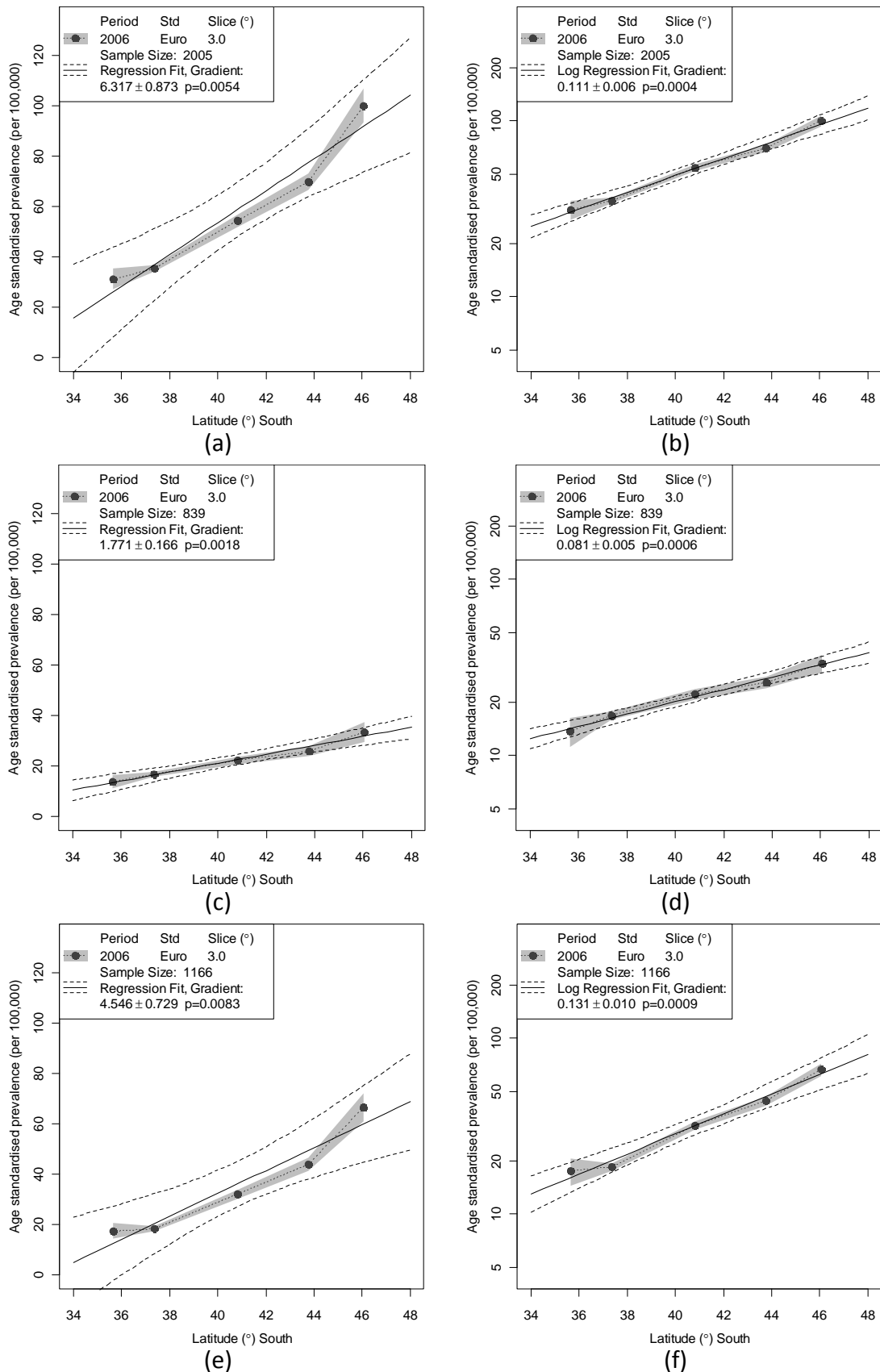


Figure 5.1 Variation in prevalence of MS in New Zealand—census day 2006
 Prevalence rates of MS in New Zealand for census day 2006, age-standardised to the European standard. **(a)** and **(b)** Combined Migrated and Resident cohorts; **(c)** and **(d)** Migrated cohort; **(e)** and **(f)** Resident cohort. Regression: **(a)**, **(c)** and **(e)** linear; **(b)**, **(d)** and **(f)** log. Latitude: population weighted centroid for 3° slices.

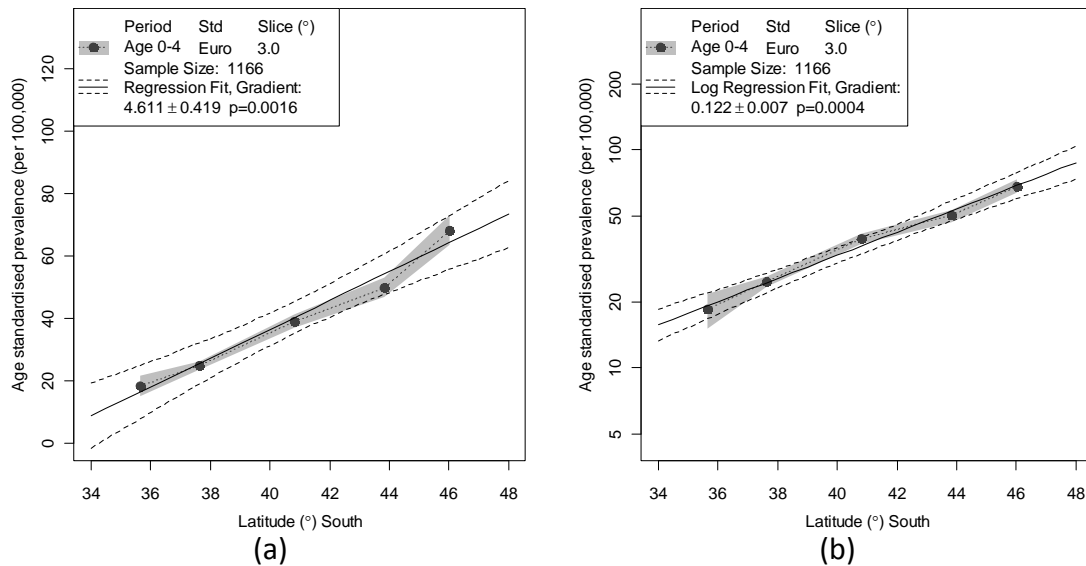


Figure 5.2 Variation in prevalence of MS in New Zealand—Age 0–4

Prevalence rates of MS in New Zealand at Age 0–4 for the Resident cohort, age-standardised to the European standard. **(a)** linear and **(b)** log regression. Latitude: population weighted centroid for 3° slices.

Cohort	Cohort Size	Period	Linear Gradient	Std Err	R^2	p -value	Anova p -value
Combined	2005	2006	6.317	0.873	0.946	0.00544	0.170
Migrated	839	2006	1.771	0.166	0.974	0.00177	0.010
Resident	1166	2006	4.546	0.729	0.928	0.00832	
Resident	1166	Age 0–4	4.611	0.419	0.976	0.00161	0.941

(a)

Cohort	Period	Log Fit Gradient	Std Err	R^2	p -value	Rate Ratio	Rate Ratio Range	Anova p -value
Combined	2006	0.1106	0.0060	0.991	3.5e-4	1.117	1.110–1.124	0.131
Migrated	2006	0.0805	0.0052	0.987	6.0e-4	1.084	1.078–1.090	0.004
Resident	2006	0.1306	0.0098	0.983	9.0e-4	1.140	1.128–1.151	
Resident	Age 0–4	0.1224	0.0067	0.991	3.7e-4	1.130	1.123–1.138	0.515

(b)

Table 5.4 Gradients of prevalence of MS with latitude in New Zealand

Regression gradients from **(a)** linear and **(b)** log regression fits to the 2006 and Age 0–4 prevalence data, age-standardised to the European standard. Cohort sizes for **(b)** are as given in **(a)**. Data is aggregated to 3° latitude slices, represented by the population weighted centroid of each slice. Rate ratios are given for the log fits. Anova p -values are for models comparing the relevant gradients with the Resident cohort 2006 gradients.

Table 5.3(d) also gives prevalence at Age 0–4 for cases resident in New Zealand from birth to onset of MS, with regression fits illustrated in Figure 5.2 and also summarised in Table 5.4. The Age 0–4 crude prevalence rates are similar to those for the full 2006 sample (Table 5.3(a)), yet the age-standardised rates are noticeably different—this is due to the single age group (0–4) and the age standardisation process. However, the rate ratio, which is obtained from the log fit, lies roughly half way between those of the two larger 2006 samples at 1.130, representing a prevalence rate increase of 13% for each degree of latitude south. The Age 0–4 and the 2006 prevalence data will be discussed further in section 5.3.2 when factors and migration will be considered.

Stratification of the prevalence data by gender and MS phenotype will be discussed in section 5.3. Before that, however, the rest of this section will look at how the choice of analysis techniques affects the results obtained. Regression techniques will be discussed in conjunction with other techniques, where appropriate, rather than in isolation. All results are for the Resident cohort, unless stated otherwise.

5.2.1 Consequences of Areal Interpolation

The effect of the size of the latitude slices used when aggregating the MS data and in the areal interpolation process for the census population can be seen in Figure 5.3, which gives data for various slice sizes from 0.2° to 4.0° for the Resident cohort. The upper graphs in Figure 5.3 give the number of MS cases per degree latitude, for each slice, and the lower graphs give the equivalent census population, after areal interpolation. The crude prevalence rate of MS (for a given slice size) is obtained by taking the ratio of the data from the appropriate upper and lower graphs. Here, the latitude taken to represent a given slice is the mean latitude of each slice. (In this section a 'slice' will refer to a latitude slice used in the areal interpolation process. In addition, 'mean latitude' will refer to the mean latitude of a slice and 'weighted latitude' will refer to the latitude of the population weighted centroid of the slice.)

As is to be expected, the larger slice sizes smooth out variations in the data. However, it is also noticeable that the variations in the distribution of the census population with latitude are mimicked quite closely by variations in the distribution of MS cases for all slice sizes, with the possible exception of the 4° slices. For this largest slice size, the averaging out of the data highlights the opposing trends of the population data, which falls slightly with increasing latitude south, and the MS data which tends to increase

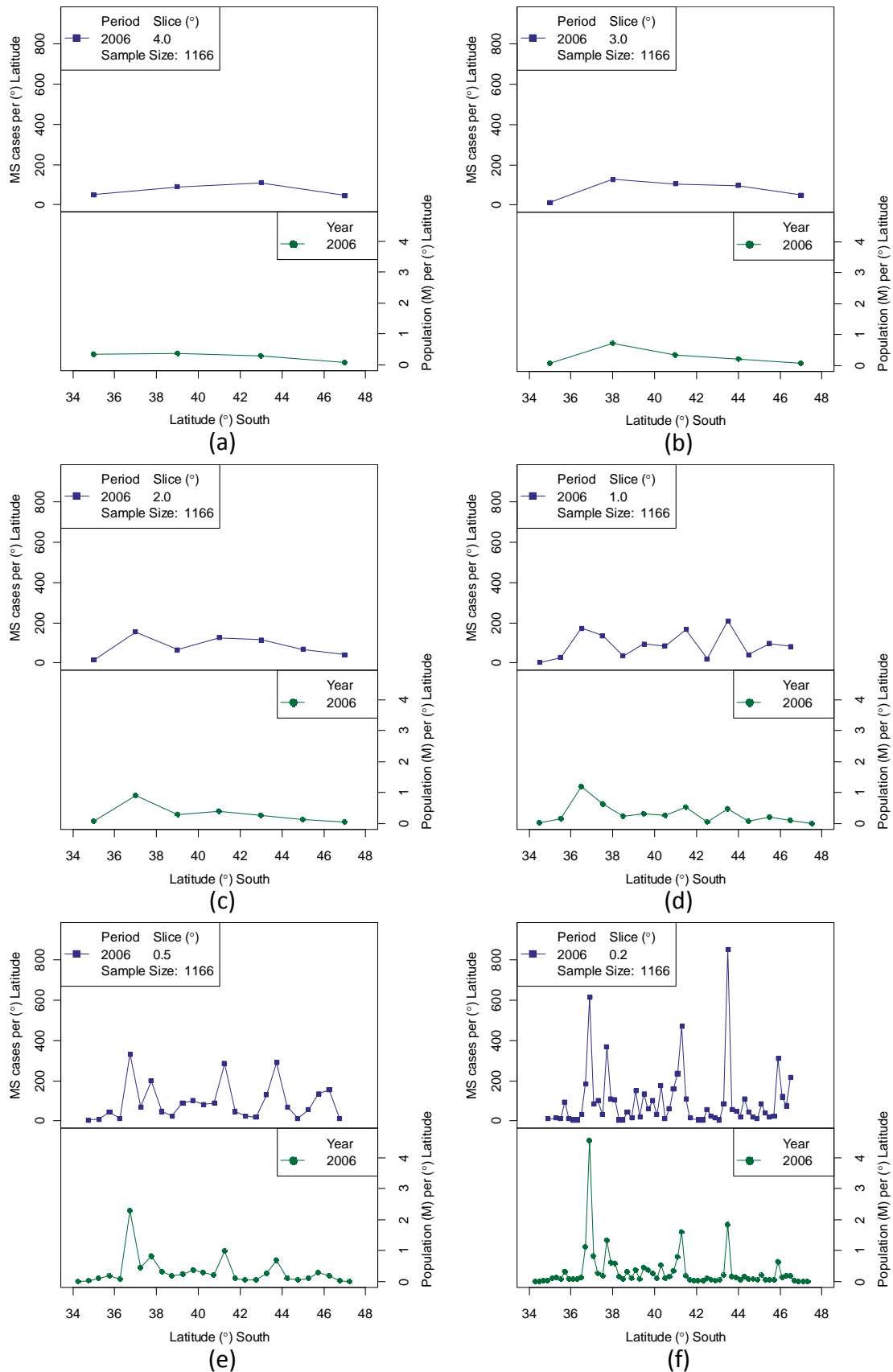


Figure 5.3 MS case numbers and population—mean latitudes

The numbers of MS cases per degree of latitude and census populations for census day 2006 with mean latitude used to represent each slice. Upper plots (blue): the number of MS cases; lower plots (green): the equivalent census populations. Latitude slices: **(a)** 4.0° **(b)** 3.0° **(c)** 2.0° **(d)** 1.0° **(e)** 0.5° **(f)** 0.2°. (Cohort: Resident.)

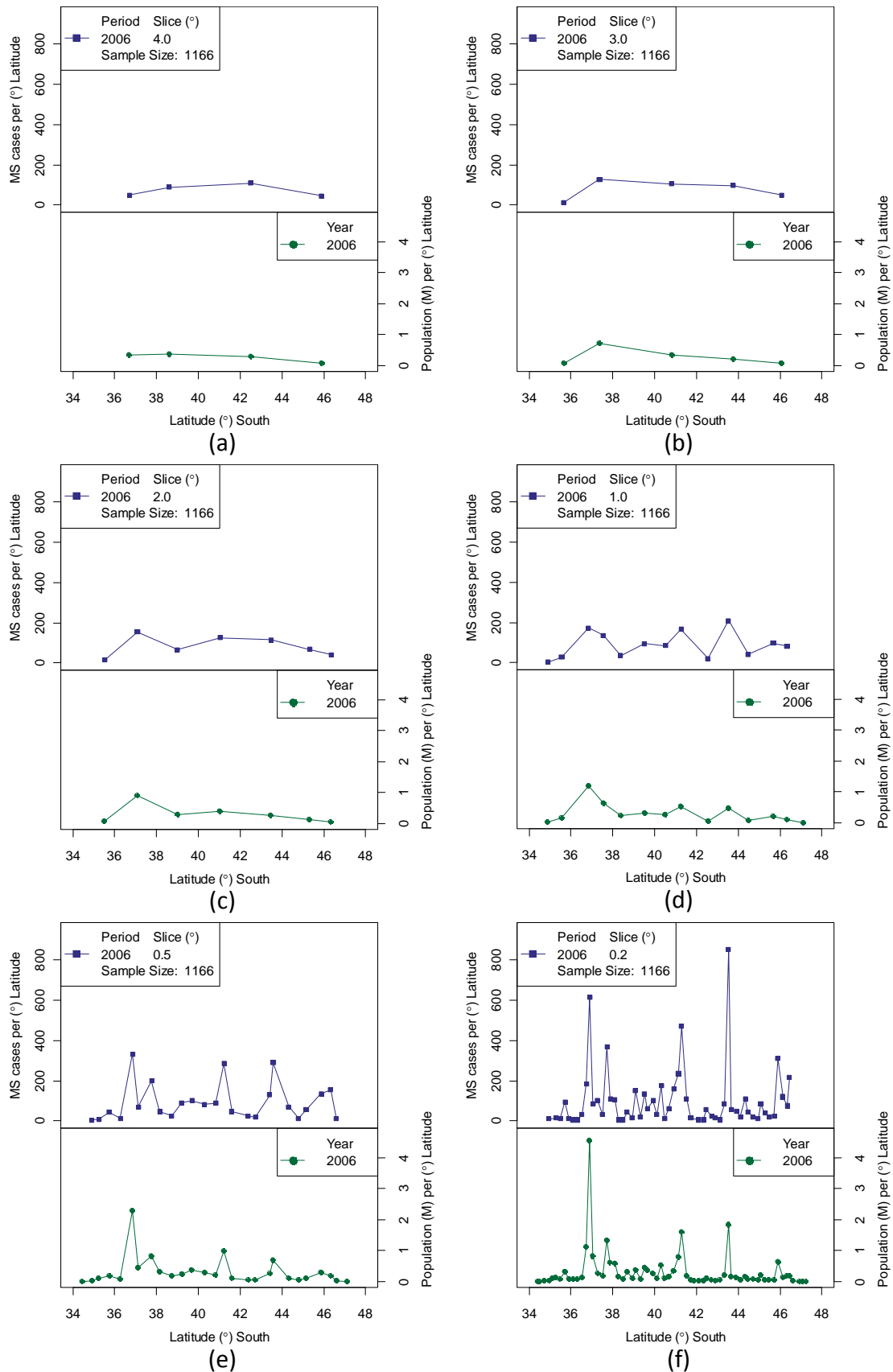


Figure 5.4 MS case numbers and population—population weighted latitudes
 The numbers of MS cases per degree of latitude and census populations for census day 2006 with the latitude of the population weighted centroid used to represent each slice. Upper plots (blue): the number of MS cases; lower plots (green): the equivalent census populations.
 Latitude slices: **(a)** 4.0° **(b)** 3.0° **(c)** 2.0° **(d)** 1.0° **(e)** 0.5° **(f)** 0.2°. (Cohort: Resident.)

with increasing latitude south. In addition, the 4° slice data lacks various peaks which are present in the data for the 3°, and smaller, slices. This is due to where the slice boundaries fall, and the representative latitude being at its mean rather than at its population weighted centroid. The data for the 0.5° and 0.2° slices indicate a very strong peak in the population at just under 37° with another smaller peak between 37.5° and 38°. These correspond to Auckland, the largest city in New Zealand and the two cities of Hamilton and Tauranga, just to the south. For the 4° slices, the majority of the Auckland City population falls into the latitude slice centred on 35° and it is averaged with the otherwise very low population of Northland (and a certain amount of sea to the north of New Zealand) which drastically reduces the population density of the slice. In contrast, the Hamilton and Tauranga populations fall in the slice centred on 39° which also contains other smaller urban areas. Figure 5.4 shows the same MS case number and census population data as in Figure 5.3, with the latitude slices now represented by the latitude of the population weighted centroid of each slice. The choice of representative latitude appears to make little difference to the general trends in the data, especially for the smaller slice sizes. However, for the 3° and 4° slices, the change in the distribution of the data points with latitude is more noticeable—the two points around Auckland, Hamilton and Tauranga, are now closer together and there is a longer gap before the third data point. The most southerly data point is also seen to be further north. These differences are sufficient to alter the gradients obtained from regression fits as will now be discussed.

The prevalence rates of MS on census day 2006, age-standardised to the European standard are shown in Figure 5.5 which gives results on a linear scale for the same slices and mean latitudes as Figure 5.3. It can clearly be seen that for the 0.5° and 0.2° slice sizes the variations in the data away from the regression lines increase with increasing prevalence rates. That is, the data is not heteroscedastic and so linear regression techniques are not strictly suitable. The equivalent log regression fits shown in Figure 5.6 illustrate the benefit of log regression—here the variations away from the regression lines are more uniform for all latitudes and so log regression fits are more suitable for this data.

In addition to a validation of the log regression technique, the log regression fits to the data indicate that the relationship between prevalence of MS cases and latitude

follows an exponential trend throughout New Zealand for census day 2006. However, comparisons of the fits for different slice sizes indicate that the rate ratio obtained varies according to the areal interpolation used for the analysis. Larger slices average out more of the natural variations in the underlying case numbers and population data, and therefore give more uniform localised rate ratios. The smaller slice sizes highlight local variations, however, it is likely that some of the more uncertain data points are being given a greater weight in the regression fits than they should have. (Although the uncertainties in the age-standardised rates are shown in the figures, these were not used in the regression calculations.) Figure 5.3 illustrates that for the sample size of 1166, and for example for the 0.2° slices, some of the slices only have a few cases, so combined with a low census population in those slices, and the process of age standardisation, the uncertainties involved when determining the prevalence rates is considerable.

Figure 5.7 and Figure 5.8 show the same 2006 prevalence data as Figure 5.5 and Figure 5.6, with the latitude of the population weighted centroid now used to represent the latitude slices rather than the mean latitude. As might be expected, this change is seen to have little or no effect for the 0.5° and 0.2° slices and greatest effect for the larger slice sizes. (The latitude of the population weighted centroid must remain within the slice it represents so can only change by up to 0.1° from the mean for the 0.2° slices, whereas it can change by up to 2° from the mean for the 4° slices.) The main effect appears to be to reduce the variation in regression fit gradients with latitude slice size. This will be discussed further, later in this section, when variations in the rate ratios are examined more closely in conjunction with the Age 0–4 data, presented next.

The prevalence rates of MS for residence location at Age 0–4 are shown in Figure 5.9 for linear regression fits and Figure 5.10 for log regression fits. The latitudes used to represent the slices are those from the population weighted centroids. (Graphs using the mean latitude are not presented here.) A visual comparison of the Age 0–4 data with the 2006 data from Figure 5.7 and Figure 5.8 shows differences in the natural variations present within each set of data, however, these differences appear to be no greater than the natural variations present. This is confirmed by the linear and log regression gradients and other fit statistics, which are summarised in Table 5.5 for mean latitudes, and Table 5.6 for weighted latitudes.

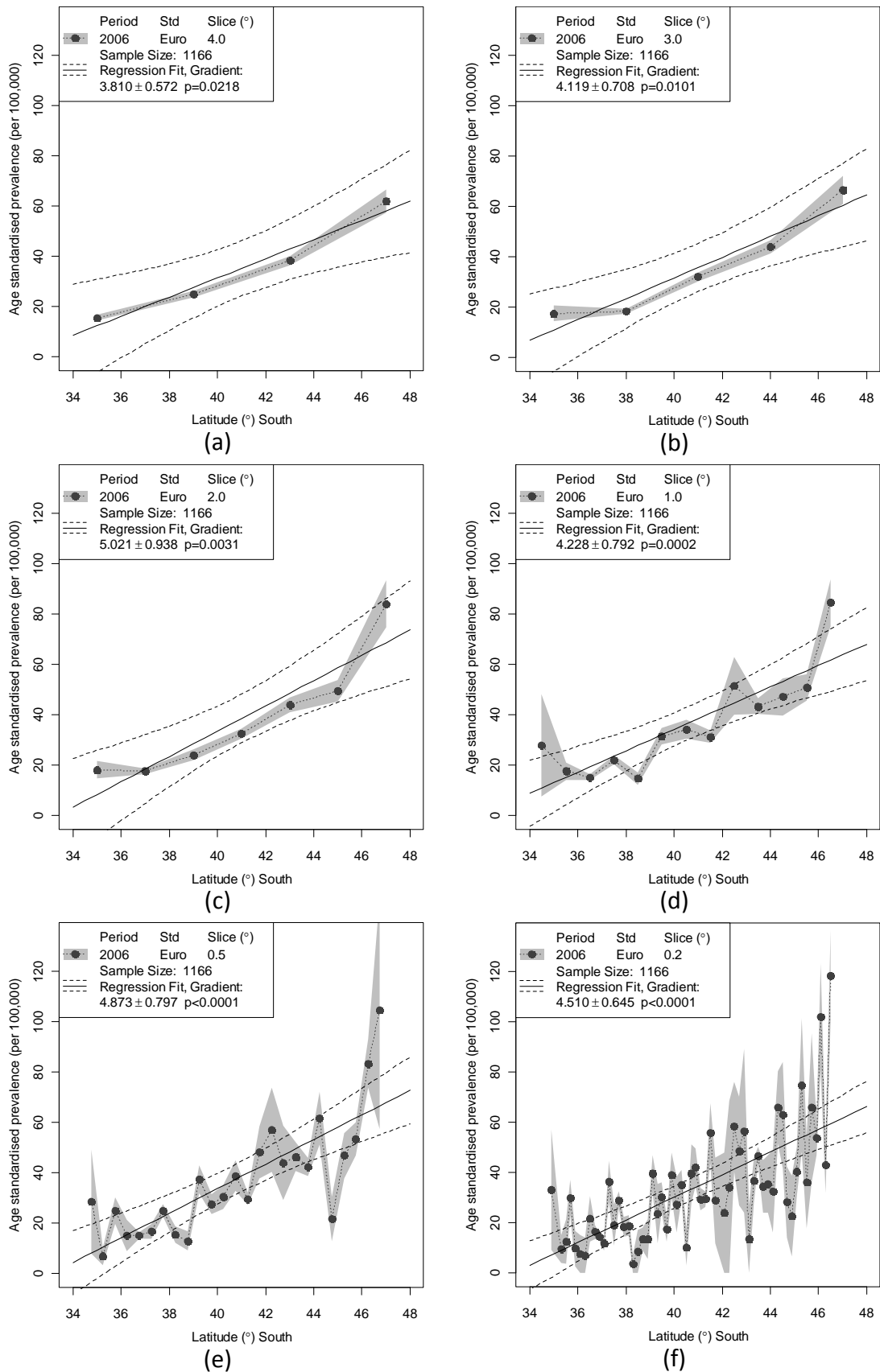


Figure 5.5 Linear regression fits—mean latitudes for 2006 Prevalence

Linear regression fits for the 2006 prevalence data standardised to the European standard, with mean latitude used to represent each slice. Shaded areas: standard error uncertainties in the data; solid lines: regression fits; dashed lines: 95% confidence bands for the regression fits.

Latitude slices: **(a)** 4.0° **(b)** 3.0° **(c)** 2.0° **(d)** 1.0° **(e)** 0.5° **(f)** 0.2°. (Cohort: Resident.)

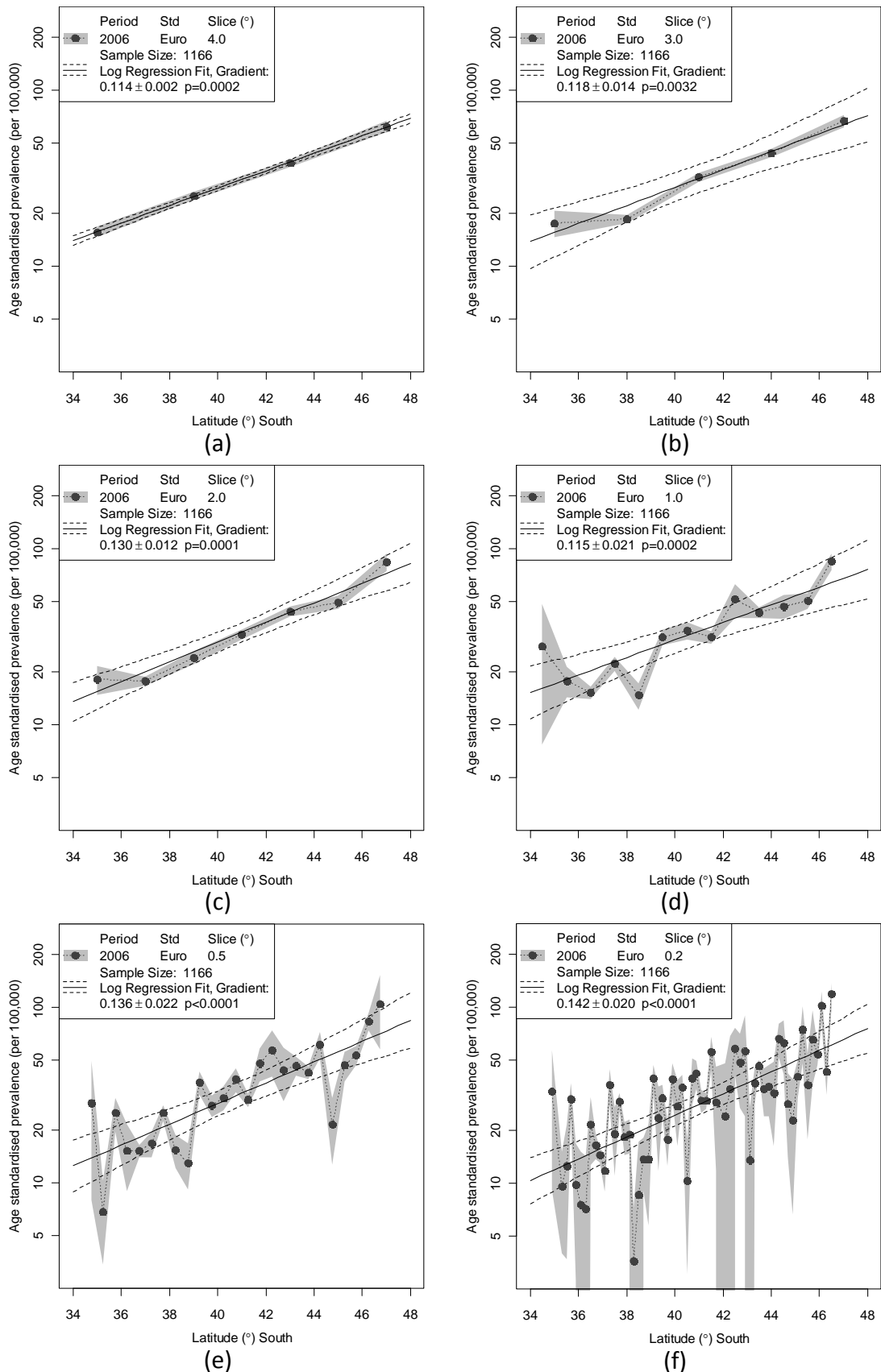


Figure 5.6 Log regression fits—mean latitudes for 2006 Prevalence

Log regression fits for the 2006 prevalence data standardised to the European standard, with mean latitude used to represent each slice. Shaded areas: standard error uncertainties in the data; solid lines: regression fits; dashed lines: 95% confidence bands for the regression fits.

Latitude slices: **(a)** 4.0° **(b)** 3.0° **(c)** 2.0° **(d)** 1.0° **(e)** 0.5° **(f)** 0.2°. (Cohort: Resident.)

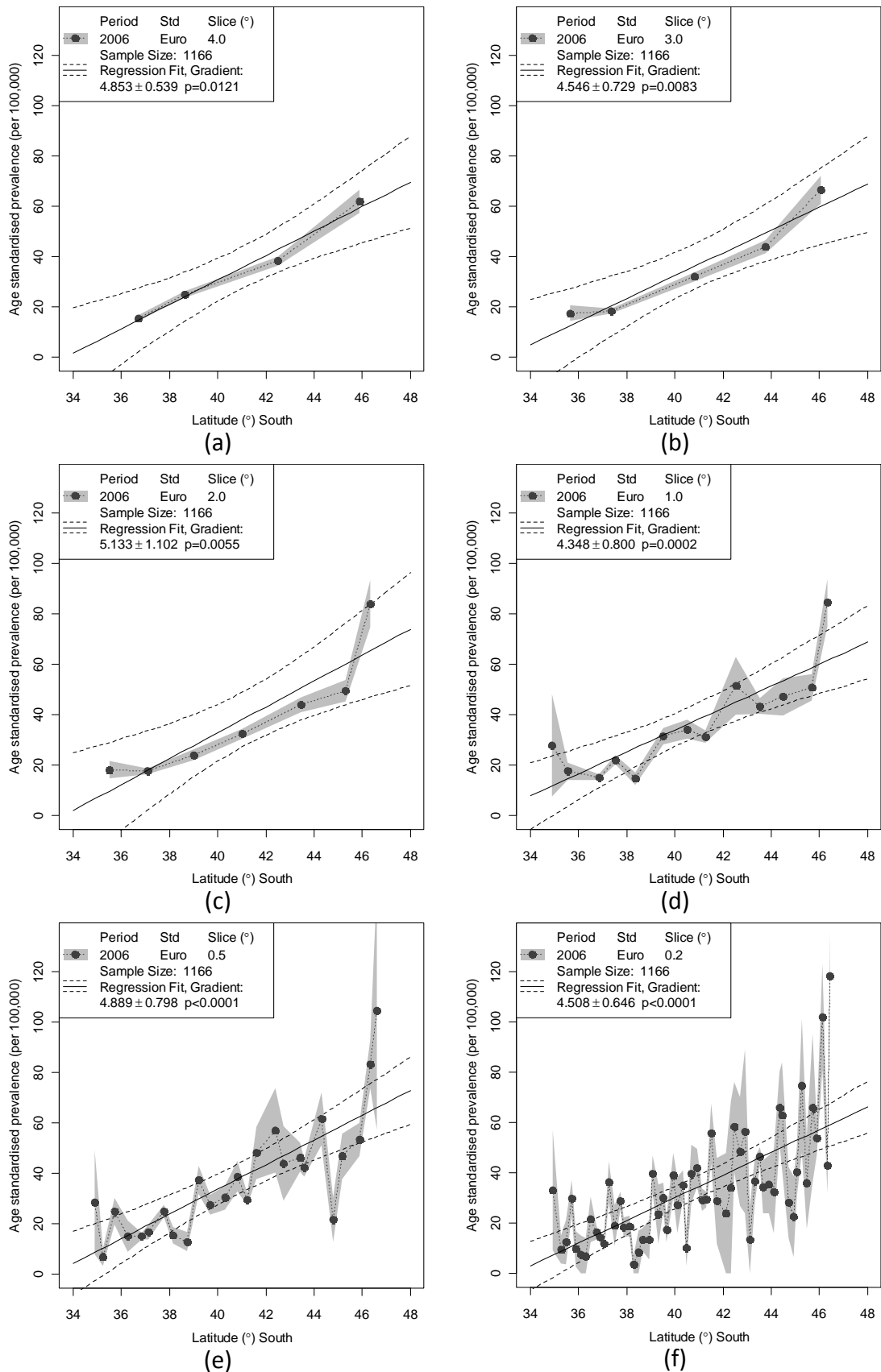


Figure 5.7 Linear regression fits—weighted latitudes for 2006 Prevalence
 Linear regression fits for the 2006 prevalence data standardised to the European standard, with the latitude of the population weighted centroid used to represent each slice. Shaded areas: standard error uncertainties in the data; solid lines: regression fits; dashed lines: 95% confidence bands for the regression fits. Latitude slices: (a) 4.0° (b) 3.0° (c) 2.0° (d) 1.0° (e) 0.5° (f) 0.2°. (Cohort: Resident.)

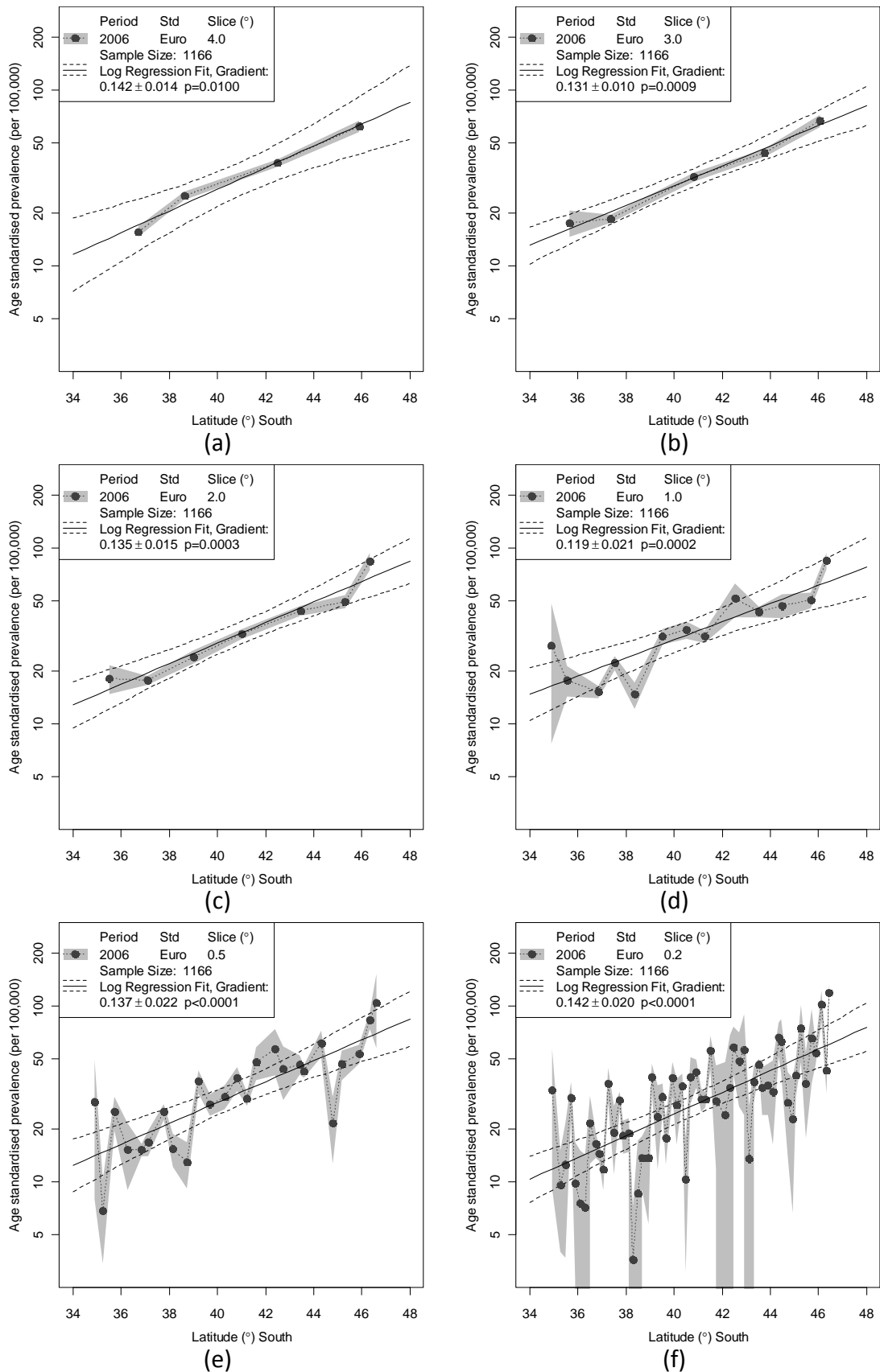


Figure 5.8 Log regression fits—weighted latitudes for 2006 Prevalence

Log regression fits for the 2006 prevalence data standardised to the European standard, with the latitude of the population weighted centroid used to represent each slice. Shaded areas: standard error uncertainties in the data; solid lines: regression fits; dashed lines: 95% confidence bands for the regression fits.

Latitude slices: **(a)** 4.0° **(b)** 3.0° **(c)** 2.0° **(d)** 1.0° **(e)** 0.5° **(f)** 0.2°. (Cohort: Resident.)

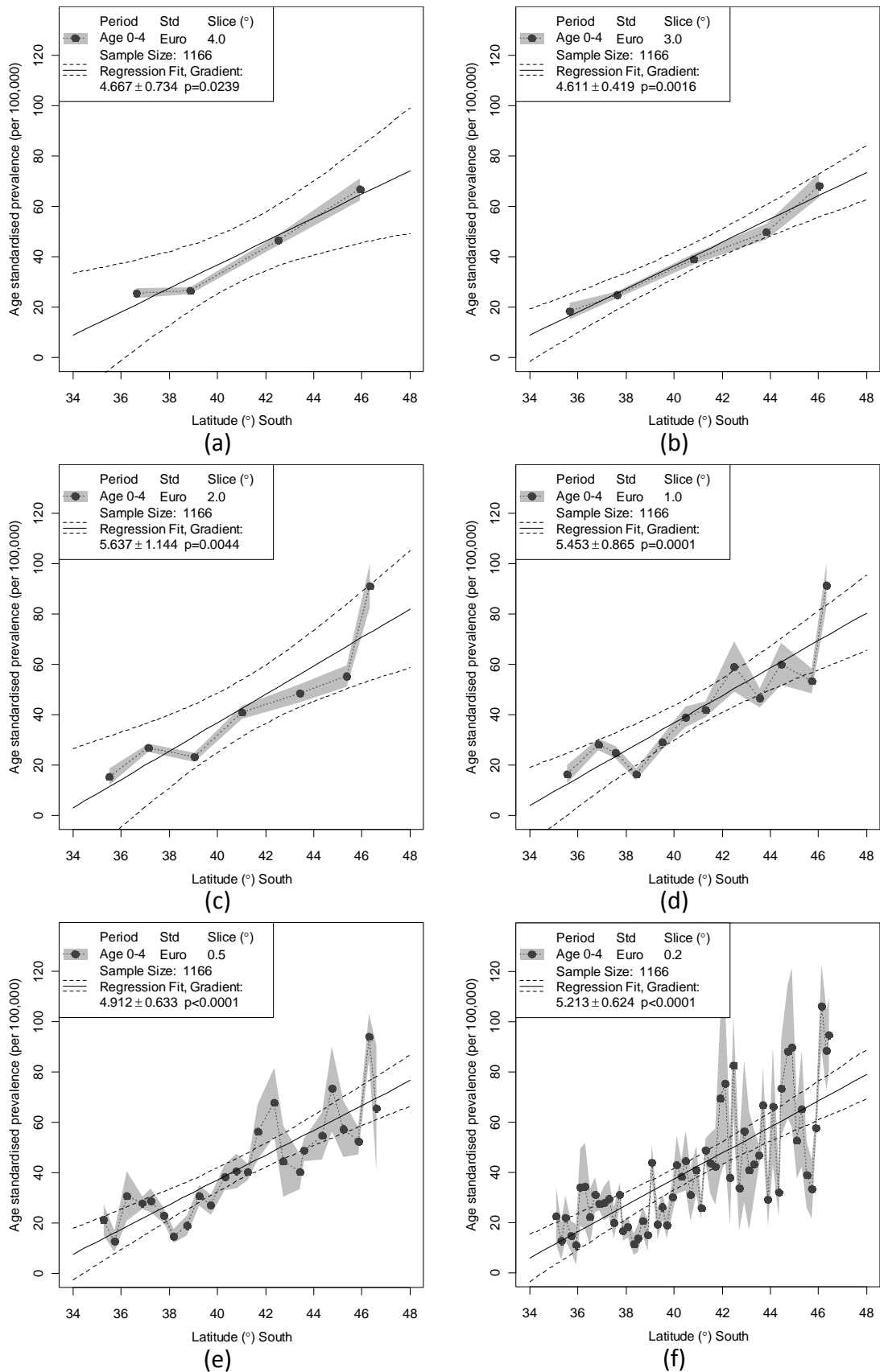


Figure 5.9 Linear regression fits—weighted latitudes for Age 0–4 Prevalence
 Linear regression fits for the Age 0–4 prevalence data standardised to the European standard, with the latitude of the population weighted centroid used to represent each slice. Shaded areas: standard error uncertainties in the data; solid lines: regression fits; dashed lines: 95% confidence bands for the regression fits. Latitude slices: **(a)** 4.0° **(b)** 3.0° **(c)** 2.0° **(d)** 1.0° **(e)** 0.5° **(f)** 0.2°. (Cohort: Resident.)

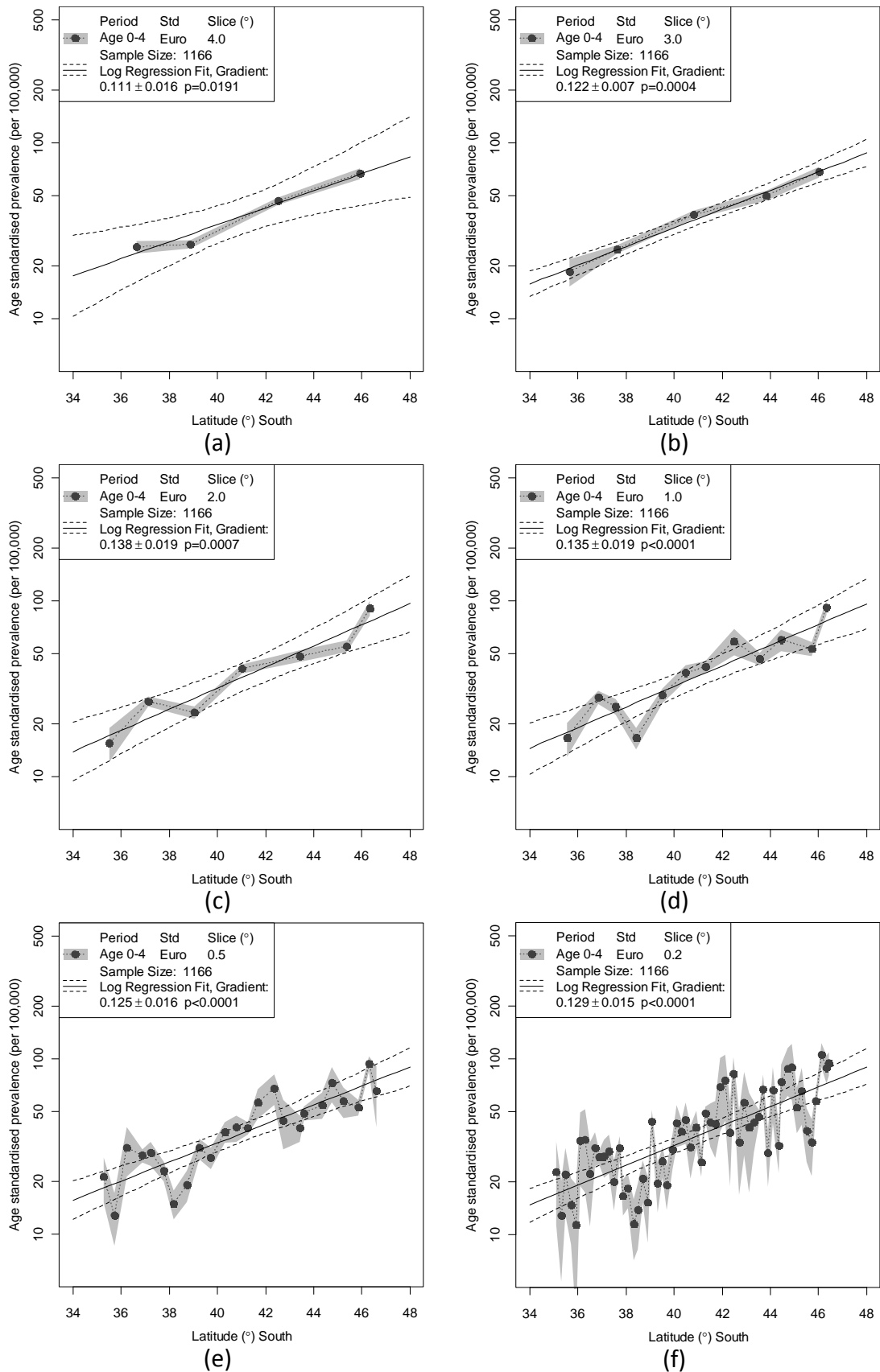


Figure 5.10 Log regression fits—weighted latitudes for Age 0–4 Prevalence
 Log regression fits for the Age 0–4 prevalence data standardised to the European standard, with the latitude of the population weighted centroid used to represent each slice. Shaded areas: standard error uncertainties in the data; solid lines: regression fits; dashed lines: 95% confidence bands for the regression fits. Latitude slices: **(a)** 4.0° **(b)** 3.0° **(c)** 2.0° **(d)** 1.0° **(e)** 0.5° **(f)** 0.2°. (Cohort: Resident.)

Latitude Slice (°)	Linear Gradient	Std Err	R^2	p -value
4.0	3.810	0.572	0.957	0.0218
3.0	4.119	0.708	0.919	0.0101
2.0	5.021	0.938	0.851	0.0031
1.0	4.228	0.792	0.721	2.4e-4
0.5	4.873	0.797	0.619	3.1e-6
0.2	4.510	0.645	0.470	4.0e-9

(a) Data: 2006; Fit: Linear

Latitude Slice (°)	Linear Gradient	Std Err	R^2	p -value
4.0	3.598	0.828	0.904	0.0491
3.0	4.154	0.367	0.977	0.0015
2.0	5.527	0.942	0.873	0.0020
1.0	5.391	0.844	0.803	8.0e-5
0.5	4.895	0.635	0.730	1.1e-7
0.2	5.217	0.622	0.557	1.8e-11

(b) Data: Age 0–4; Fit: Linear

Latitude Slice (°)	Log Fit Gradient	Std Err	R^2	p -value	Rate Ratio	Rate Ratio Range
4.0	0.1140	0.0017	1.000	2.3e-4	1.121	1.119–1.123
3.0	0.1176	0.0136	0.962	0.0032	1.125	1.110–1.140
2.0	0.1296	0.0123	0.957	1.3e-4	1.138	1.124–1.152
1.0	0.1151	0.0210	0.732	1.9e-4	1.122	1.100–1.146
0.5	0.1361	0.0218	0.629	2.3e-6	1.146	1.121–1.171
0.2	0.1420	0.0200	0.477	2.8e-9	1.153	1.130–1.176

(c) Data: 2006; Fit: Log

Latitude Slice (°)	Log Fit Gradient	Std Err	R^2	p -value	Rate Ratio	Rate Ratio Range
4.0	0.0861	0.0181	0.919	0.0415	1.089	1.070–1.110
3.0	0.1102	0.0058	0.992	3.2e-4	1.117	1.110–1.123
2.0	0.1339	0.0151	0.940	3.0e-4	1.143	1.126–1.161
1.0	0.1330	0.0192	0.828	4.0e-5	1.142	1.121–1.164
0.5	0.1244	0.0157	0.740	7.2e-8	1.133	1.115–1.150
0.2	0.1294	0.0146	0.583	3.1e-12	1.138	1.122–1.155

(d) Data: Age 0–4; Fit: Log

Table 5.5 Linear and log regression gradients—mean latitudes

Gradients from linear and log regression fits to the 2006 and Age 0–4 prevalence data with mean latitude used to represent each slice. Data is age-standardised to the European standard. Rate ratios are given for the log fits. (Cohort: Resident.)

Latitude Slice (°)	Linear Gradient	Std Err	R^2	p -value
4.0	4.853	0.539	0.976	0.0121
3.0	4.546	0.729	0.928	0.0083
2.0	5.133	1.102	0.813	0.0055
1.0	4.348	0.800	0.729	2.1e-4
0.5	4.889	0.798	0.620	3.0e-6
0.2	4.508	0.646	0.469	4.2e-9

(a) Data: 2006; Fit: Linear

Latitude Slice (°)	Linear Gradient	Std Err	R^2	p -value
4.0	4.667	0.734	0.953	0.0239
3.0	4.611	0.419	0.976	0.0016
2.0	5.637	1.144	0.829	0.0044
1.0	5.453	0.865	0.799	8.9e-5
0.5	4.912	0.633	0.732	9.7e-8
0.2	5.213	0.624	0.555	2.0e-11

(b) Data: Age 0–4; Fit: Linear

Latitude Slice (°)	Log Fit Gradient	Std Err	R^2	p -value	Rate Ratio	Rate Ratio Range
4.0	0.1424	0.0143	0.980	0.0100	1.153	1.137–1.170
3.0	0.1306	0.0098	0.983	9.0e-4	1.140	1.128–1.151
2.0	0.1346	0.0148	0.943	2.7e-4	1.144	1.127–1.161
1.0	0.1187	0.0210	0.743	1.5e-4	1.126	1.103–1.150
0.5	0.1368	0.0218	0.632	2.0e-6	1.147	1.122–1.172
0.2	0.1422	0.0201	0.477	2.7e-9	1.153	1.130–1.176

(c) Data: 2006; Fit: Log

Latitude Slice (°)	Log Fit Gradient	Std Err	R^2	p -value	Rate Ratio	Rate Ratio Range
4.0	0.1113	0.0156	0.962	0.0191	1.118	1.100–1.135
3.0	0.1224	0.0067	0.991	3.7e-4	1.130	1.123–1.138
2.0	0.1383	0.0187	0.917	7.0e-4	1.148	1.127–1.170
1.0	0.1351	0.0193	0.831	3.7e-5	1.145	1.123–1.167
0.5	0.1249	0.0157	0.743	6.2e-8	1.133	1.115–1.151
0.2	0.1294	0.0146	0.582	3.3e-12	1.138	1.122–1.155

(d) Data: Age 0–4; Fit: Log

Table 5.6 Linear and log regression gradients—weighted latitudes

Gradients from linear and log regression fits to the 2006 and Age 0–4 prevalence data with the latitude of the population weighted centroid used to represent each slice. Data is age-standardised to the European standard. Rate ratios are given for the log fits. (Cohort: Resident.)

The rate ratios obtained from the fit gradients for the different slice sizes are also shown in Figure 5.11. Note that, apart from the rate ratios for the 4° slices, the prevalence rate ratios for 2006 and for Age 0–4 agree within the uncertainties of the standard errors for each of the slice sizes. However, for the 1° and 2° slice sizes the Age 0–4 ratios are larger than the 2006 rate ratios, whereas for the other slice sizes, the 2006 rate ratios are the largest. There is also a noticeable effect due to the choice of latitude slice size for the mean latitude fits—the two largest latitude slice sizes (with fewest data points) give lower rate ratios than the smaller latitude slice sizes (which have a greater number of data points). For the weighted latitude fits, there is a suggestion of a similar trend with the Age 0–4 data, however, it is much less pronounced. Since the use of the weighted latitudes appears to limit the effect of the latitude slice size, weighted latitudes from the population weighted centroids will be used in further analyses rather than the mean latitudes.

From Figure 5.11, it is not obvious whether any particular choice of latitude slice size should be preferred over any other. It is, however, clear that any rate ratio results obtained are likely to depend (to some extent) upon the choice of latitude slice size. As mentioned previously in this section, the choice of boundaries for the latitude slices is also likely to affect the rate ratios obtained from the analyses, though this was not investigated further in this study. The summary results presented at the start of this

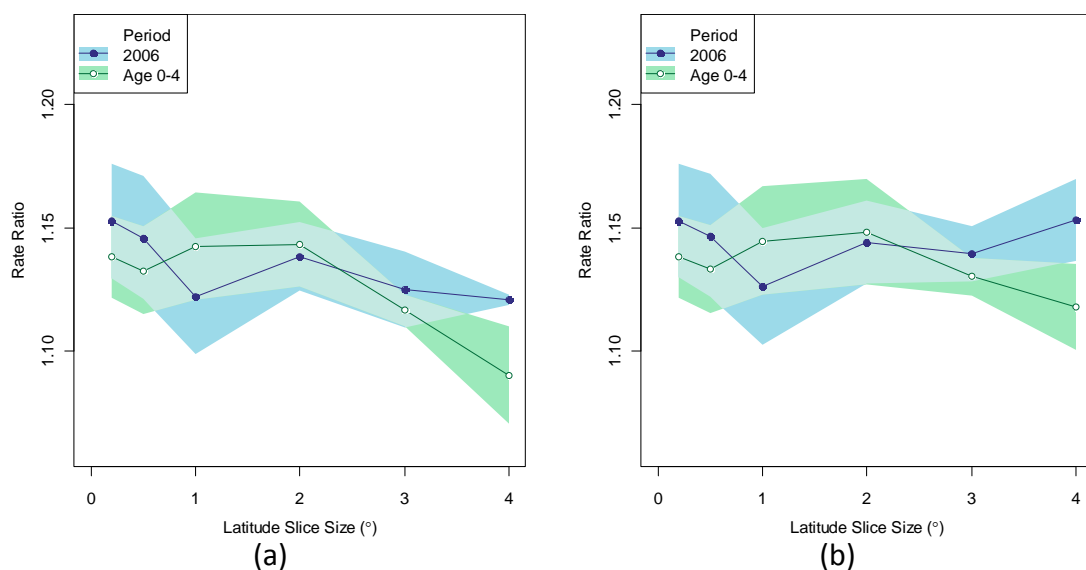


Figure 5.11 Rate Ratios—Latitude Slice Size

MS prevalence rate ratios from log regression fits for prevalence on census day 2006 (blue) and Age 0–4 (green). Shaded areas: 95% confidence bands for the rate ratios as determined from the log regression fits. Latitude slices: **(a)** mean latitude; **(b)** latitude of the population weighted centroid. (Cohort: Resident.)

section were based on 3° slices since that size appears to give the best fits when results are taken as a whole (including results presented later in this chapter). However, rather than choose one particular slice size for use in further analyses, the choice will be made based on other criteria, as will now be explained. The data presented so far in this section is for all MS cases considered. Once the data is stratified (section 5.3) and the potential number of cases in each slice decreases, it would be prudent to use a larger slice size rather than a smaller size in order to reduce potential uncertainties. However, it must also be remembered that use of a single slice size may not be appropriate—for example, simply from Figure 5.11(b), a comparison of the rate ratios for the 2006 prevalence data and the Age 0–4 data based on 2° slices would give a completely different interpretation from a comparison based on 4° slices. Therefore, results for 2°, 3° and 4° slices will be presented in most of section 5.3, with a simple north-south divide used to improve the sample numbers even further when looking at month of birth effects in section 5.3.4.

The original cohort size of 2917 reported by Taylor et al (2010b) was reduced to 1166 cases for the Resident cohort, as mentioned in section 4.2.1. 912 cases were omitted due to lack of questionnaire data, and 839 (42% of the remaining cases) were in a separate Migrated cohort due to the strict residence criterion. If a less strict residence criterion had been used, allowing some of the 839 Migrated cases to remain in the Resident cohort, then some of the fluctuations and uncertainties which are apparent in the results presented here may have been reduced. However, this would have been at the cost of introducing possible confounding effects due to migration into or out of New Zealand.

Further discussion of the Age 0–4 data and the 2006 prevalence data in connection with various factors and the effects of migration are in section 5.3.2. Before continuing to section 5.3, however, this overview section (5.2) is concluded with a discussion of results looking at the effects of the choice of standard population in the age standardisation process.

5.2.2 Choice of Standard for Age Standardisation

The various regression fits carried out for section 5.2.1 using the European standard population (and weighted latitudes) were repeated using the Segi standard population. The Segi standard (Figure 4.7) has a different age structure which was designed to be

more representative of a World population, with higher low age group rates and lower high age group rates, than the Scandinavian based European standard. (The term 'standard' will refer in this section to standard population.) The results for the linear regression and log regression gradients are given in Table 5.7. Although the linear regression results differ from those presented in Table 5.6, the gradients from the log regression fits are almost identical for the 2006 prevalence data, and identical for the Age 0–4 data. For the Age 0–4 data only one age group from the census and standard populations is required so the change of standard amounts to a fixed scale factor in the calculation of age-standardised prevalence, and thus gives the same gradients in the log regression fits, as described in section 4.3.3. Similarly, for the 2006 prevalence data, the change in standard would only affect the gradient of the log regression fits if the age structure of the cases varied with latitude. Figure 5.12 shows the age group profile with latitude for 4° slices from which it can be seen that there are some variations, however, the general profile is similar for all slices. This leads to little change in the age-standardised rates and hence the gradients obtained from the log regression fits. (Age profiles for other slice sizes are similar, however, for clarity, only the 4° slice profiles are illustrated.)

Since the choice of standard has little effect on the log regression gradients and rate ratios for this data, the European standard population will be used for all further results presented here.

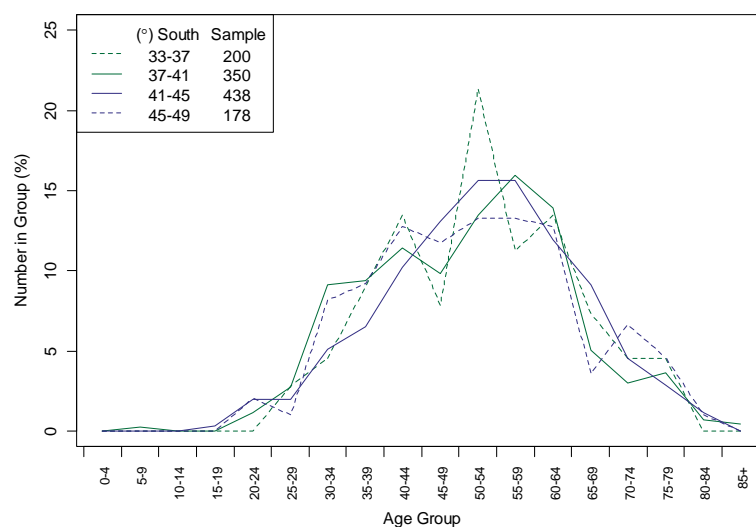


Figure 5.12 Variation of age structure with Latitude

The proportions of MS cases in each age group on census day 2006 for 4° latitude slices, normalised by the number of cases in each slice (Sample). (Cohort: Resident.)

Latitude Slice (°)	Linear Gradient	Std Err	R^2	p -value
4.0	3.793	0.398	0.978	0.0109
3.0	3.540	0.558	0.931	0.0079
2.0	4.021	0.870	0.810	0.0057
1.0	3.459	0.618	0.740	1.6E-04
0.5	3.872	0.642	0.613	3.8E-06
0.2	3.517	0.514	0.460	6.9E-09

(a) Data: 2006; Fit: Linear

Latitude Slice (°)	Linear Gradient	Std Err	R^2	p -value
4.0	7.001	1.101	0.953	0.0239
3.0	6.916	0.628	0.976	0.0016
2.0	8.455	1.717	0.829	0.0044
1.0	8.179	1.297	0.799	8.9E-05
0.5	7.368	0.949	0.732	9.7E-08
0.2	7.820	0.935	0.555	2.0E-11

(b) Data: Age 0–4; Fit: Linear

Latitude Slice (°)	Log Fit Gradient	Std Err	R^2	p -value	Rate Ratio	Rate Ratio Range
4.0	0.1430	0.0148	0.979	0.0105	1.154	1.137–1.171
3.0	0.1309	0.0094	0.985	8.2E-04	1.140	1.129–1.151
2.0	0.1352	0.0148	0.944	2.6E-04	1.145	1.128–1.162
1.0	0.1212	0.0200	0.769	8.3E-05	1.129	1.106–1.152
0.5	0.1377	0.0219	0.633	2.0E-06	1.148	1.123–1.173
0.2	0.1400	0.0203	0.465	5.3E-09	1.150	1.127–1.174

(c) Data: 2006; Fit: Log

Latitude Slice (°)	Log Fit Gradient	Std Err	R^2	p -value	Rate Ratio	Rate Ratio Range
4.0	0.1113	0.0156	0.962	0.0191	1.118	1.100–1.135
3.0	0.1224	0.0067	0.991	3.7E-04	1.130	1.123–1.138
2.0	0.1383	0.0187	0.917	7.0E-04	1.148	1.127–1.170
1.0	0.1351	0.0193	0.831	3.7E-05	1.145	1.123–1.167
0.5	0.1249	0.0157	0.743	6.2E-08	1.133	1.115–1.151
0.2	0.1294	0.0146	0.582	3.3E-12	1.138	1.122–1.155

(d) Data: Age 0–4; Fit: Log

Table 5.7 Linear and log regression gradients—Segi Standard

Gradients from linear and log regression fits to the 2006 and Age 0–4 prevalence data with the latitude of the population weighted centroid used to represent each slice. Data is age-standardised to the Segi standard. Rate ratios are given for the log fits. (Cohort: Resident.)

5.3 Effect of Factors on Prevalence

It is well established, as discussed in Chapter 2, that many factors affect the prevalence of MS. In this section, results will now be presented which examine stratification of the data by gender and MS phenotype, with a look at the effects of migration. Since stratification of the data reduces the numbers of cases in each category, latitude slice sizes of less than 2° will not be considered. In addition, the latitude slices will be represented by the latitude of the population weighted centroid (PWC) and data will be age-standardised using the European standard population. The effect of month of birth on prevalence rates will also be considered, though sample numbers require some changes in the methods used, as will be discussed.

5.3.1 Gender and MS Phenotype

The number of cases of MS varies with both gender and MS phenotype. For the cohort of cases considered in this study, the overall numbers are given in Table 5.8, where cases for the RRMS and SPMS phenotypes are grouped together, as discussed in Chapter 2. The female to male (F:M) ratio for all MS phenotypes is 3.3:1 which compares with: 2.35:1 in Sweden (Ahlgren et al. 2011); 2.17:1 in Finland (Krökki et al. 2011); 3.2:1 in Canada (Orton et al. 2006); and 4.2:1 in First Nations Peoples, Alberta, Canada (Svenson et al. 2007). Note also that this ratio has been changing with time, as discussed in Chapter 2, and reviewed by Sellner et al (2011). From Sellner et al, this change would appear to be due to a difference in MS phenotype since the F:M ratio among RRMS cases has increased during the last 2–6 decades, yet the ratio among PPMS cases has changed little. Here, from Table 5.8, it can be seen that the F:M ratio

	All	RR/SPMS	PPMS	% RR/SPMS	% PPMS	RR/SP:PP
All	1166	955	204	81.9	17.5	4.68:1
Female	894	765	124	85.6	13.9	6.17:1
Male	272	190	80	69.9	29.4	2.38:1
% Female	76.7	80.1	60.8			
% Male	23.3	19.9	39.2			
F:M	3.29:1	4.03:1	1.55:1			

Table 5.8 Numbers of cases of MS by Gender and MS Phenotype

The number of cases of MS resident in New Zealand from birth to onset of MS. F:M : female:male ratio; RR/SP:PP : RR/SPMS:PPMS ratio. (The 'All' phenotype also includes seven cases where phenotype was not RRMS, SPMS or PPMS.)

does depend on MS phenotype: within RR/SPMS cases the F:M ratio rises to as high as 4:1, and for PPMS cases it drops to around 1.5:1. Therefore, a direct comparison of overall F:M rates with other studies may not be appropriate without also considering other factors including the times of the studies and the MS phenotypes of the cases.

It should be noted that the summary figures given in Table 5.8 differ slightly from those of the original study of Taylor et al (2010b) given in Table 3.7. These differences can be attributed to the proportion of male cases omitted due to incomplete data (27%) and the strict residence criterion (26%), both figures being slightly higher than the original proportion of male cases (25%).

Looking at MS phenotype by gender, just over 85% of female cases are diagnosed with RR/SPMS, compared with just under 70% of male cases, though both genders are still more likely to be diagnosed with RR/SPMS than PPMS. Variations in the prevalence–latitude rate ratios with gender and MS phenotype will now be examined.

Rather than present graphs of the data on both a linear scale, with linear regression fits, and a log scale, with log regression fits, only the log regression fits will be shown, though the linear regression fit gradients will be given in the relevant tables. Note also that comparisons between log regression fits are made based on the p -value resulting from the Anova test on the gradients, not on the rate ratios which are determined from the gradients. Where reference is made to a statistical comparison of rate ratios, this will imply a comparison of the relevant log regression gradients.

Log regression fits are shown by gender and latitude in Figure 5.13 for the 2006 prevalence data and in Figure 5.14 for the Age 0–4 data. The fit gradients and other fit statistics are also given in Table 5.9, and the rate ratios are shown in Figure 5.15 for the different latitude slice sizes.

[This section continues after the following pages of results.]

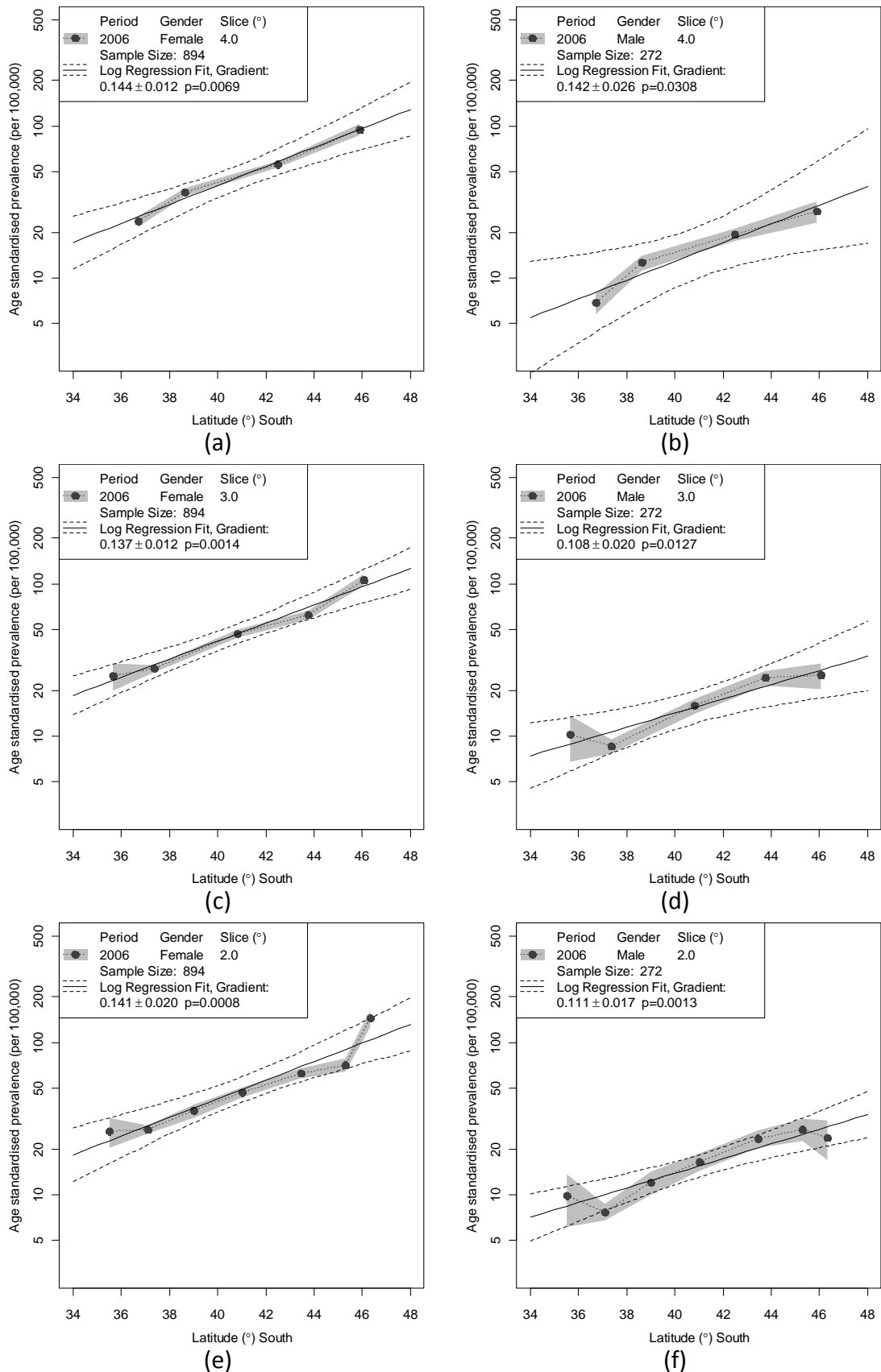


Figure 5.13 Log regression fits—2006 Prevalence by Gender

Log regression fits for the 2006 prevalence data for female and male cases, and all MS phenotypes. Shaded areas: standard error uncertainties in the data; solid lines: regression fits; dashed lines: 95% confidence bands for the regression fits.

Gender: **(a) (c) (e)** Female, **(b) (d) (f)** Male; Latitude slices: **(a) (b)** 4.0° **(c) (d)** 3.0° **(e) (f)** 2.0°. (Standard: Euro; Latitude: PWC; Cohort: Resident.)

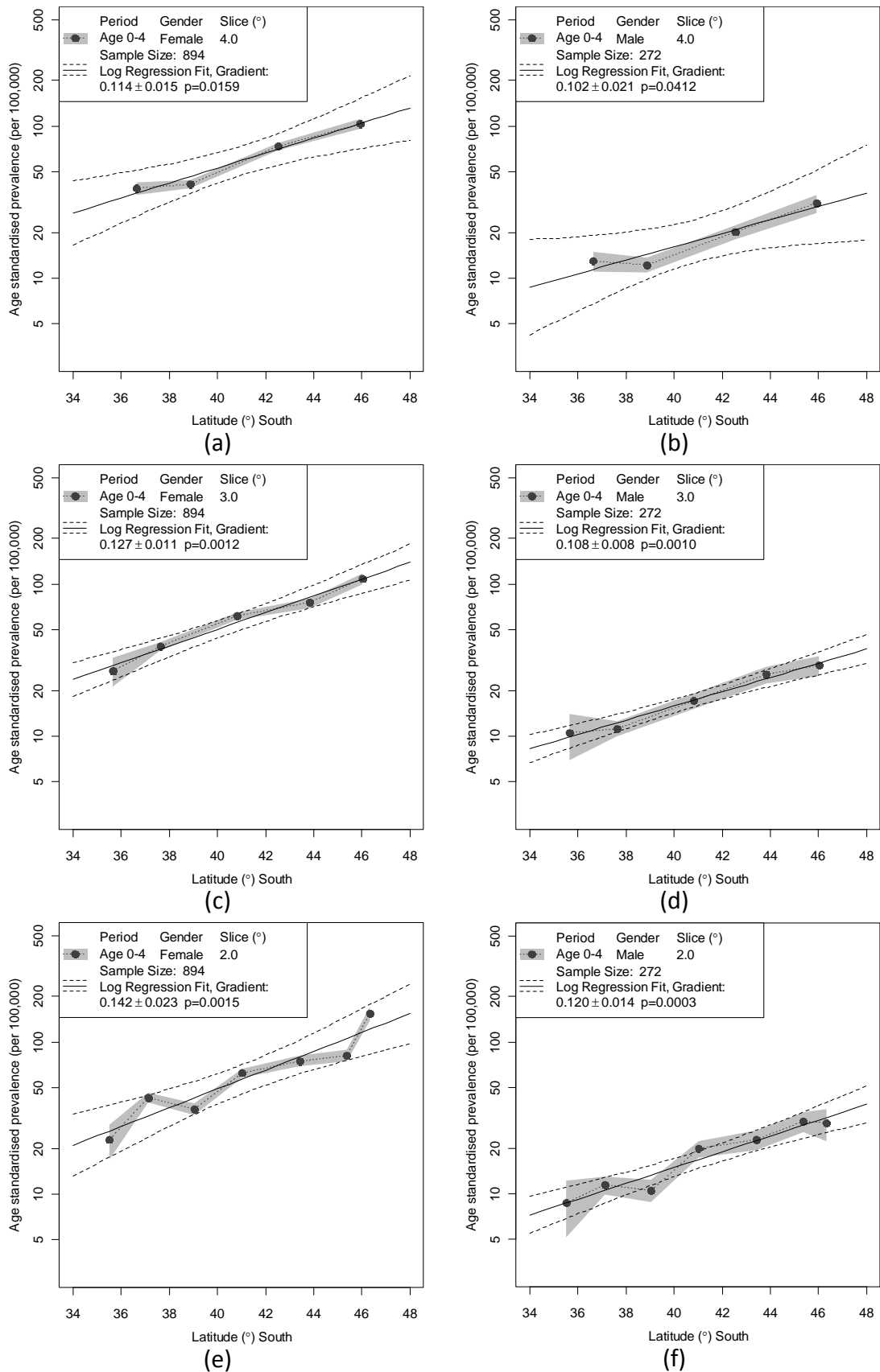


Figure 5.14 Log regression fits—Age 0–4 Prevalence by Gender

Log regression fits for the Age 0–4 prevalence data for female and male cases, and all MS phenotypes. Shaded areas: standard error uncertainties in the data; solid lines: regression fits; dashed lines: 95% confidence bands for the regression fits.

Gender: **(a) (c) (e)** Female, **(b) (d) (f)** Male; Latitude slices: **(a) (b)** 4.0° **(c) (d)** 3.0° **(e) (f)** 2.0°. (Standard: Euro; Latitude: PWC; Cohort: Resident.)

Latitude Slice (°)	Gender	Linear Regression Fit				Log Regression Fit						Anova <i>p</i> -value	
		Gradient	Std Err	R^2	<i>p</i> -value	Gradient	Std Err	R^2	<i>p</i> -value	Rate Ratio	Rate Ratio Range	Male – Female	
4.0	Female	7.489	1.029	0.964	0.0184	0.144	0.012	0.986	0.0069	1.155	1.141–1.169	0.946	
	Male	2.181	0.135	0.992	0.0038	0.142	0.026	0.939	0.0308	1.153	1.124–1.183		
3.0	Female	7.268	1.485	0.889	0.0163	0.137	0.012	0.978	0.0014	1.147	1.133–1.161	0.254	
	Male	1.712	0.277	0.927	0.0085	0.108	0.020	0.905	0.0127	1.114	1.091–1.136		
2.0	Female	8.514	2.370	0.721	0.0157	0.141	0.020	0.911	8.4e-4	1.151	1.129–1.174	0.280	
	Male	1.759	0.246	0.911	8.4e-4	0.111	0.017	0.893	0.0013	1.117	1.098–1.137		

(a) Regression fits to the 2006 data

Latitude Slice (°)	Gender	Linear Regression Fit				Log Regression Fit						Anova <i>p</i> -value	
		Gradient	Std Err	R^2	<i>p</i> -value	Gradient	Std Err	R^2	<i>p</i> -value	Rate Ratio	Rate Ratio Range	Male – Female	2006 – Age 0–4
4.0	Female	7.365	1.048	0.961	0.0197	0.114	0.015	0.968	0.0159	1.120	1.104–1.137	0.682	0.181
	Male	2.046	0.473	0.903	0.0495	0.102	0.021	0.919	0.0412	1.108	1.084–1.132		
3.0	Female	7.372	0.843	0.962	0.0032	0.127	0.011	0.980	0.0012	1.135	1.123–1.147	0.209	0.962
	Male	1.944	0.196	0.970	0.0022	0.108	0.008	0.982	0.0010	1.114	1.105–1.123		
2.0	Female	9.242	2.280	0.767	0.0098	0.142	0.023	0.889	0.0015	1.153	1.127–1.179	0.414	0.989
	Male	2.100	0.228	0.944	2.5e-4	0.120	0.014	0.939	3.1e-4	1.127	1.112–1.143		

(b) Regression fits to the Age 0–4 data

Table 5.9 Linear and log regression gradients—Gender

Gradients from linear and log regression fits to the female and male, 2006 and Age 0–4 prevalence data. Rate ratios are given for the log fits.

Anova *p*-values are for models comparing the female and male log gradients, and the 2006 and Age 0–4 log gradients.

Cohort: Resident; Sample sizes: 894 Female; 272 Male.

The first impression is that the rate ratios for the female cases are in general greater than those for the male cases. However, the differences are not statistically significant and the uncertainties involved are as great as any of the differences. In addition, further examination of the regression fits suggests that uncertainties in the extreme north and south data points could be sufficient to affect the gradients of the regression fits—omitting these points from the data would probably provide regression fits with higher correlation coefficients, lower uncertainties and closer agreement of results for different latitude slice sizes. (Note, however, that this conclusion is based solely on a visual examination of the graphs, no fits with reduced data sets were actually carried

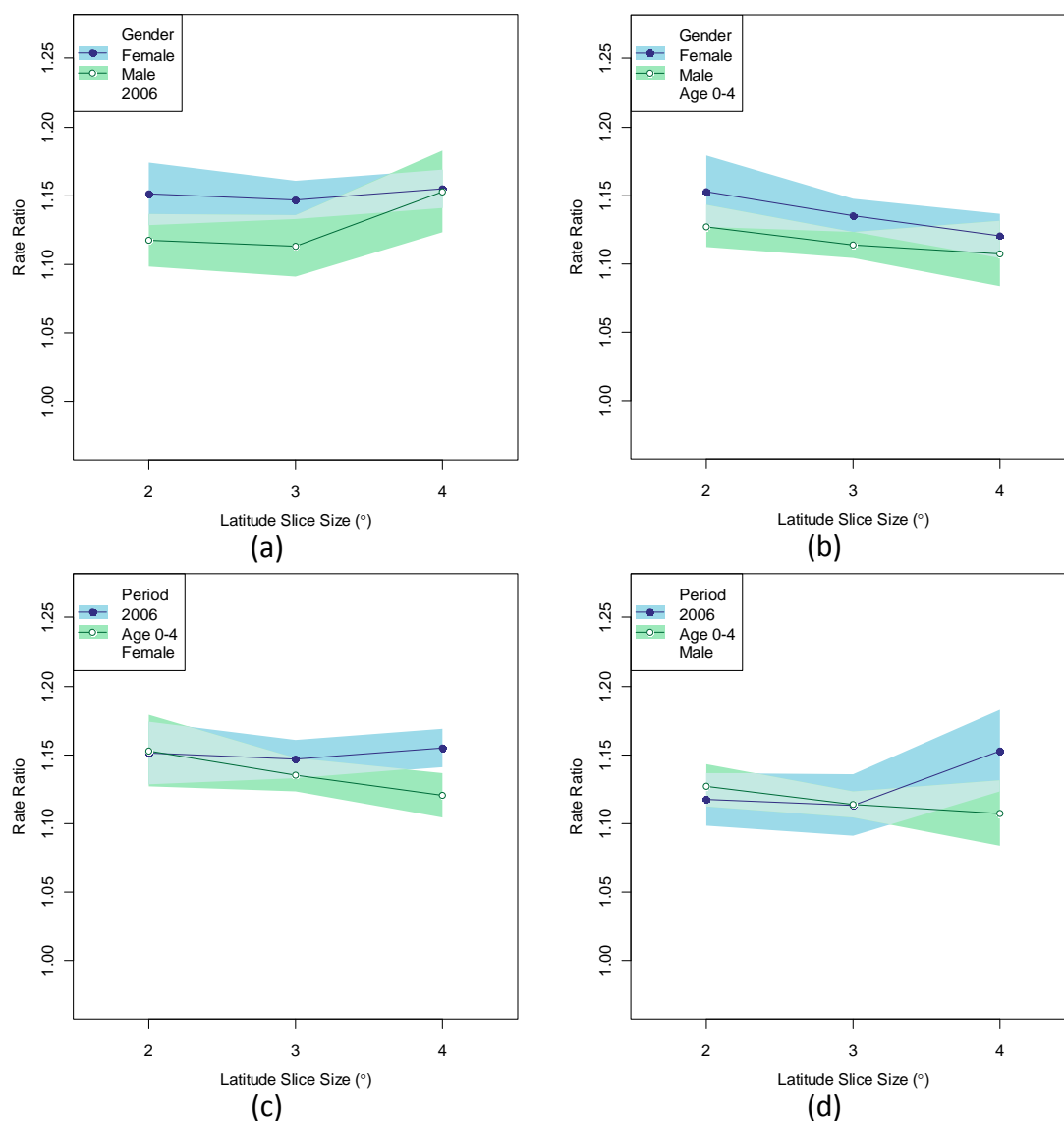


Figure 5.15 Rate Ratios—Gender and Life Period

MS prevalence rate ratios from log regression fits. Shaded areas: 95% confidence bands for the rate ratios. **(a) (b)** Rate ratios for female and male cases for: **(a)** 2006; **(b)** Age 0–4. **(c) (d)** Rate ratios for 2006 and Age 0–4 for: **(c)** female cases; **(d)** male cases. (Standard: Euro; Latitude: PWC; Cohort: Resident.)

out.) It is also apparent that any differences between the 2006 prevalence rate ratios and the Age 0–4 rate ratios are not statistically significant for either the female or the male cases.

Although there is no apparent difference in the rate ratios obtained from the data upon stratification by gender, these rate ratio analyses do not take into account the MS phenotypes of the cases. Further stratification of the data by MS phenotype will now be considered. If the rate ratios vary by gender and phenotype then this could suggest that whatever causes any latitude effect acts differently for the different gender–phenotype combinations. Table 5.8 also shows that stratification of the data by both gender and MS phenotype reduces the number of cases in each category by different degrees—there are nearly ten times more female RR/SPMS cases than there are male PPMS cases. This must be kept in mind when looking at the latitude gradients for the stratified data, since the lower numbers introduce greater uncertainties in the prevalence data. (However, these uncertainties have not been used in the regression fits reported here.)

Log regression fits for the 2006 prevalence data for RR/SPMS and PPMS phenotypes are shown in Figure 5.16 for female cases with the Age 0–4 data shown in Figure 5.17. The fit gradients and other fit statistics are given in Table 5.10. The log regression fits for male cases are shown in Figure 5.18 for the 2006 data and Figure 5.19 for the Age 0–4 data, with the fit gradients and other fit statistics given in Table 5.11. Note that the Anova p -value results for comparisons between the 2006 and Age 0–4 log regression fits are given in Table 5.10(a) for female cases and Table 5.11(b) for male cases, and comparisons between the female and male log regression fits are given in Table 5.10(b) for Age 0–4 prevalence data and Table 5.11(a) for 2006 prevalence data.

[This section continues after the following pages of results.]

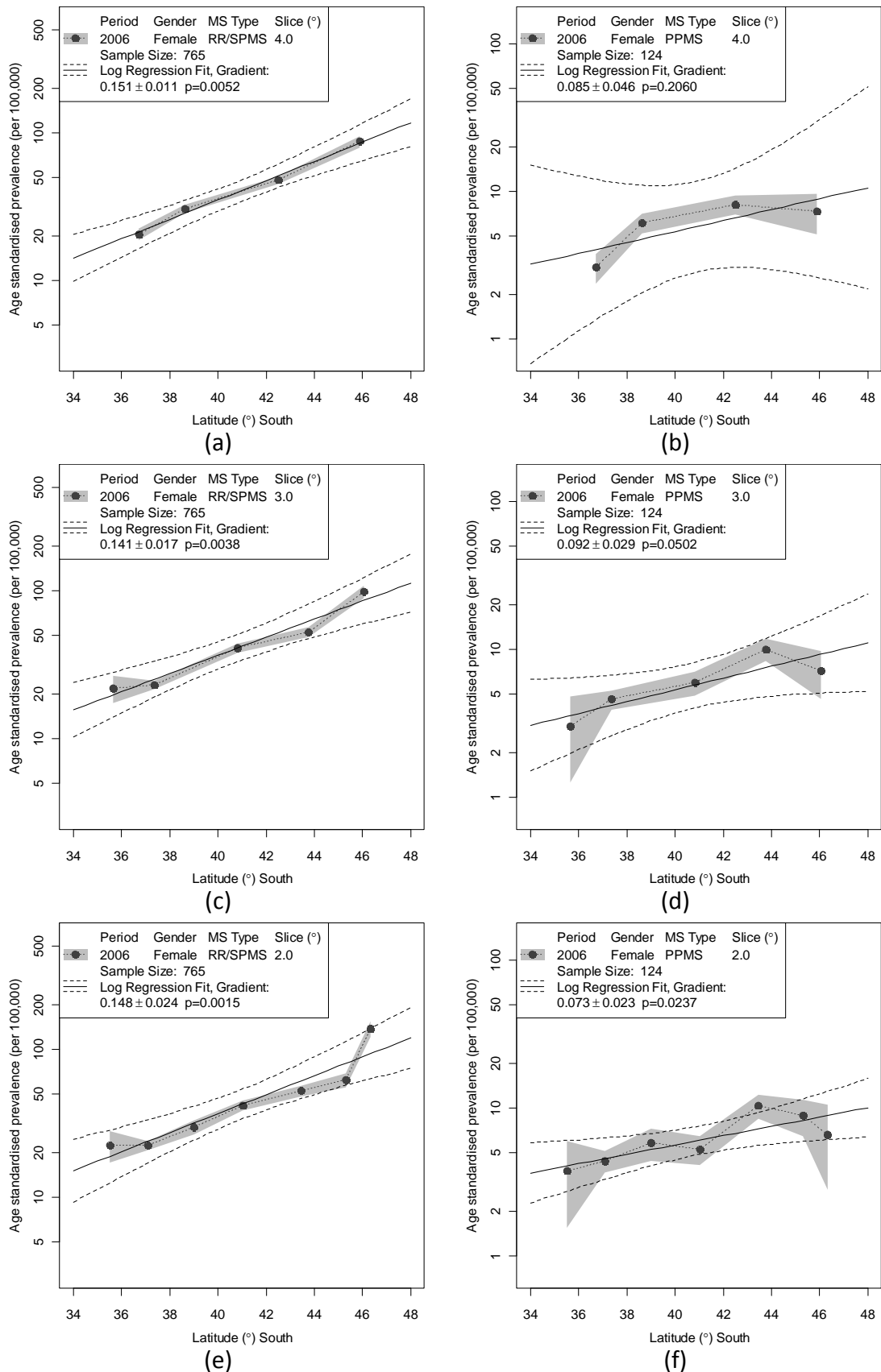


Figure 5.16 Log regression fits—2006 Prevalence by MS Phenotype (Female)
 Log regression fits for the 2006 prevalence data for female cases, by MS phenotype. Shaded areas: standard error uncertainties in the data; solid lines: regression fits; dashed lines: 95% confidence bands for the regression fits. MS phenotype: **(a) (c) (e)** RR/SPMS **(b) (d) (f)** PPMS; Latitude slices: **(a) (b)** 4.0° **(c) (d)** 3.0° **(e) (f)** 2.0°. (Standard: Euro; Latitude: PWC; Cohort: Resident.)

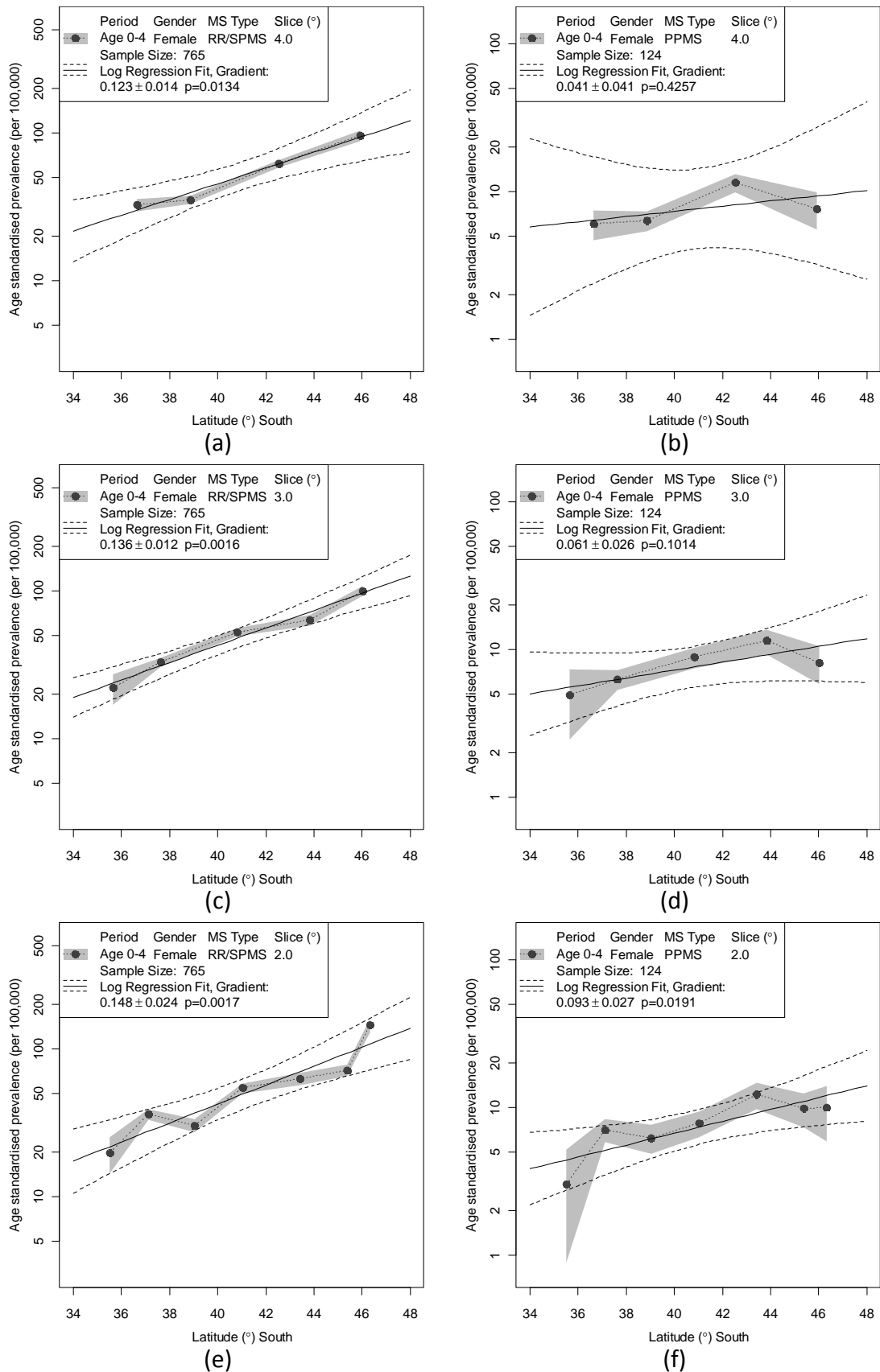


Figure 5.17 Log regression fits—Age 0–4 Prevalence by MS Phenotype (Female)
 Log regression fits for the Age 0–4 prevalence data for female cases, by MS phenotype. Shaded areas: standard error uncertainties in the data; solid lines: regression fits; dashed lines: 95% confidence bands for the regression fits. MS phenotype: **(a) (c) (e)** RR/SPMS **(b) (d) (f)** PPMS; Latitude slices: **(a) (b)** 4.0° **(c) (d)** 3.0° **(e) (f)** 2.0°. (Standard: Euro; Latitude: PWC; Cohort: Resident.)

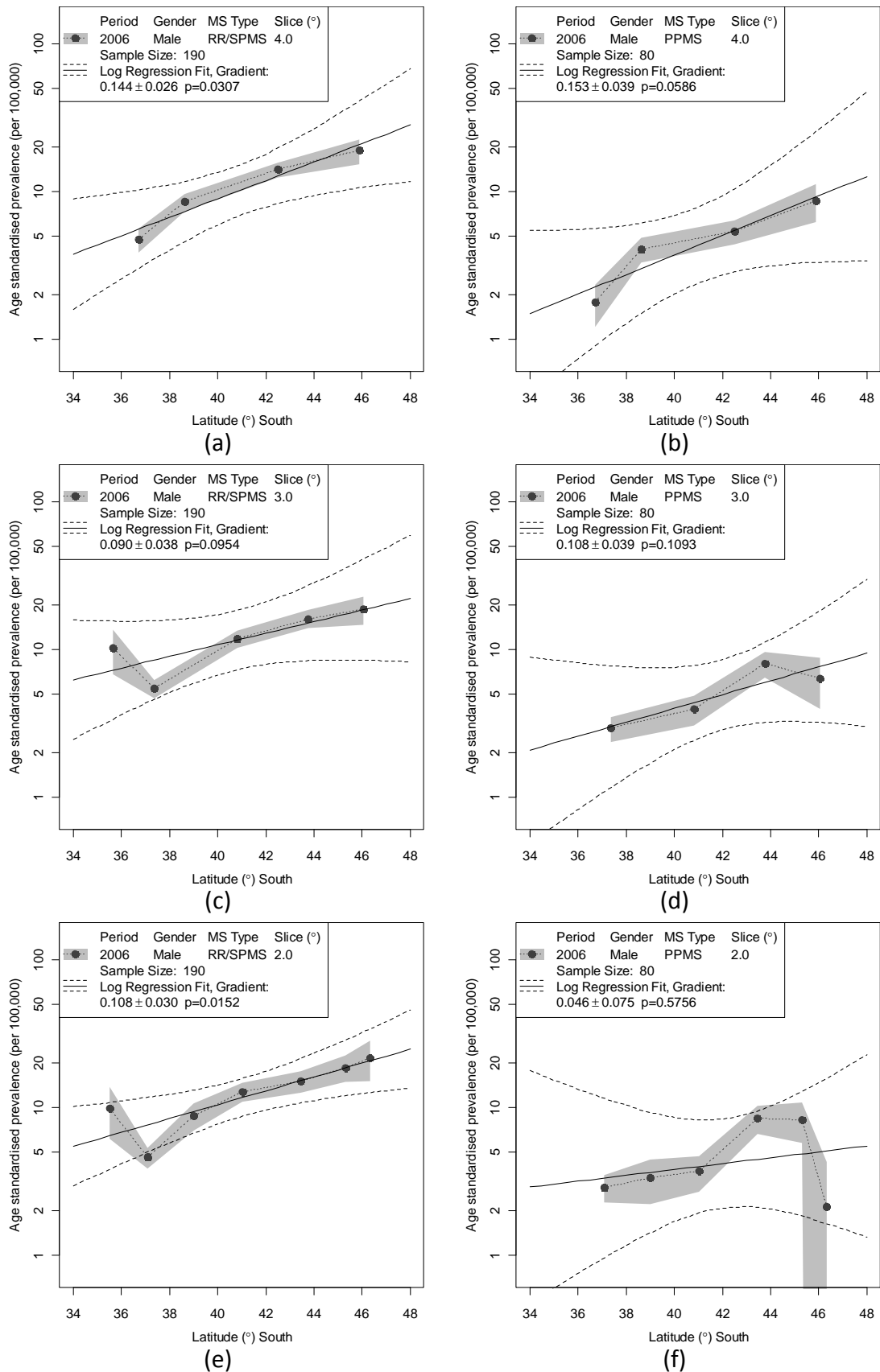


Figure 5.18 Log regression fits—2006 Prevalence by MS Phenotype (Male)
 Log regression fits for the 2006 prevalence data for male cases, by MS phenotype. Shaded areas: standard error uncertainties in the data; solid lines: regression fits; dashed lines: 95% confidence bands for the regression fits. MS phenotype: **(a) (c) (e)** RR/SPMS **(b) (d) (f)** PPMS; Latitude slices: **(a) (b)** 4.0° **(c) (d)** 3.0° **(e) (f)** 2.0°. (Standard: Euro; Latitude: PWC; Cohort: Resident.)

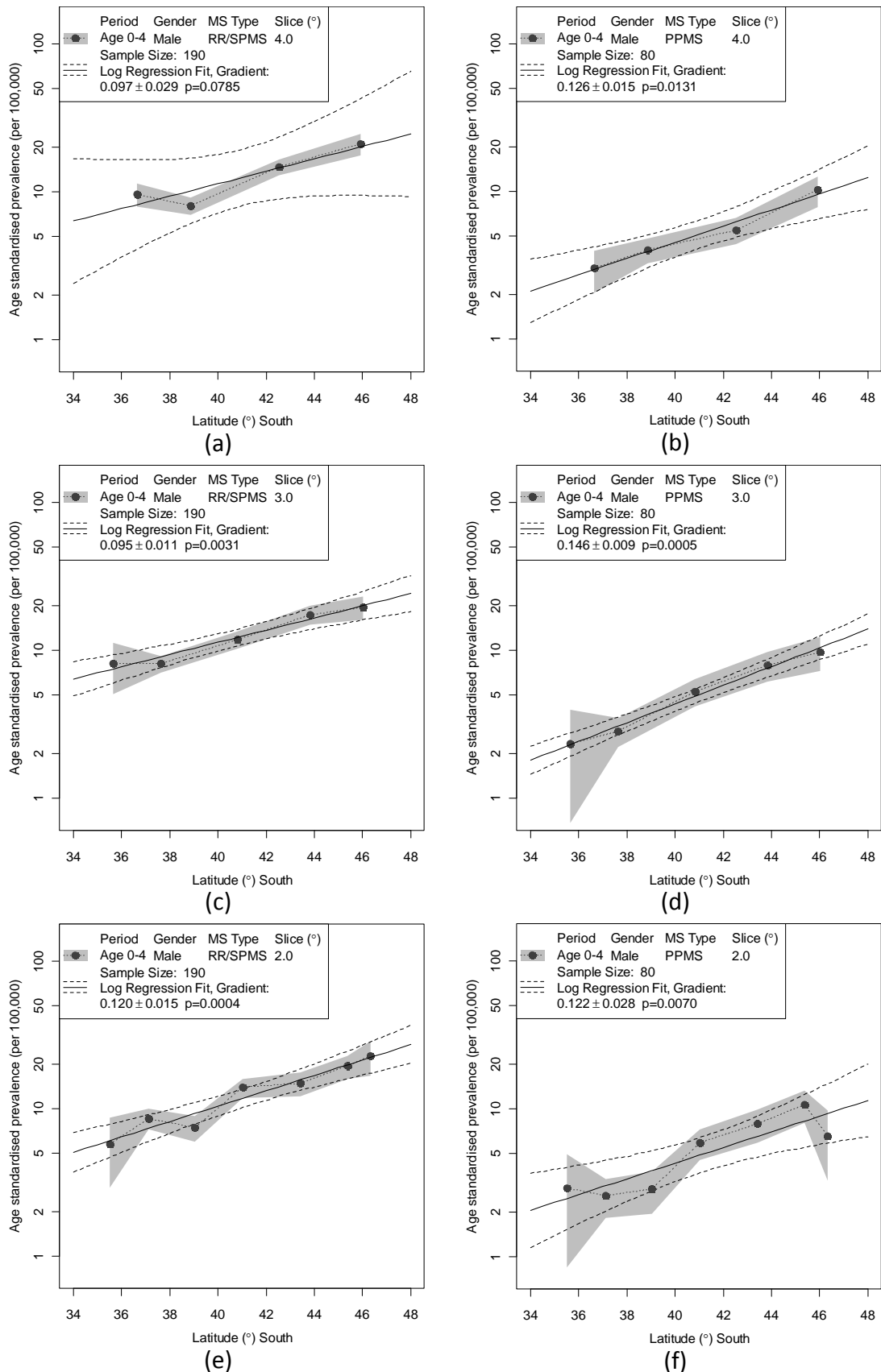


Figure 5.19 Log regression fits—Age 0–4 Prevalence by MS Phenotype (Male)
 Log regression fits for the Age 0–4 prevalence data for male cases, by MS phenotype. Shaded areas: standard error uncertainties in the data; solid lines: regression fits; dashed lines: 95% confidence bands for the regression fits. MS phenotype: **(a) (c) (e)** RR/SPMS **(b) (d) (f)** PPMS; Latitude slices: **(a) (b)** 4.0° **(c) (d)** 3.0° **(e) (f)** 2.0°. (Standard: Euro; Latitude: PWC; Cohort: Resident.)

Latitude Slice (°)	MS Phenotype	Linear Regression Fit				Log Regression Fit						Anova <i>p</i> -value	
		Gradient	Std Err	R^2	<i>p</i> -value	Gradient	Std Err	R^2	<i>p</i> -value	Rate Ratio	Rate Ratio Range	RR/SPMS – PPMS	2006 – Age 0–4
4.0	RR/SPMS	6.971	1.203	0.944	0.0285	0.151	0.011	0.990	0.0052	1.163	1.150–1.176	0.241	0.192
	PPMS	0.443	0.225	0.660	0.1874	0.085	0.046	0.630	0.2060	1.089	1.040–1.141		0.800
3.0	RR/SPMS	6.681	1.639	0.847	0.0266	0.141	0.017	0.957	0.0038	1.152	1.132–1.172	0.193	0.989
	PPMS	0.512	0.197	0.691	0.0810	0.092	0.029	0.771	0.0502	1.096	1.065–1.128		0.509
2.0	RR/SPMS	8.035	2.513	0.672	0.0240	0.148	0.024	0.887	0.0015	1.159	1.132–1.187	0.045	0.468
	PPMS	0.443	0.170	0.576	0.0478	0.073	0.023	0.674	0.0237	1.075	1.051–1.100		0.584

(a) Regression fits to the female 2006 data

Latitude Slice (°)	MS Phenotype	Linear Regression Fit				Log Regression Fit						Anova <i>p</i> -value	
		Gradient	Std Err	R^2	<i>p</i> -value	Gradient	Std Err	R^2	<i>p</i> -value	Rate Ratio	Rate Ratio Range	RR/SPMS – PPMS	Male – Female
4.0	RR/SPMS	7.003	1.167	0.947	0.0267	0.123	0.014	0.973	0.0134	1.130	1.114–1.147	0.132	0.468
	PPMS	0.311	0.368	0.263	0.4871	0.041	0.041	0.330	0.4257	1.041	1.000–1.085		0.121
3.0	RR/SPMS	6.863	1.054	0.934	0.0074	0.136	0.012	0.976	0.0016	1.145	1.131–1.159	0.043	0.047
	PPMS	0.449	0.221	0.580	0.1345	0.061	0.026	0.646	0.1014	1.063	1.036–1.092		0.023
2.0	RR/SPMS	8.595	2.352	0.728	0.0147	0.148	0.024	0.883	0.0017	1.160	1.132–1.188	0.158	0.343
	PPMS	0.617	0.170	0.724	0.0152	0.093	0.027	0.699	0.0191	1.097	1.068–1.127		0.463

(b) Regression fits to the female Age 0–4 data

Table 5.10 Linear and log regression gradients—MS Phenotype (Female)

Gradients from linear and log regression fits to the female RR/SPMS and PPMS phenotypes, 2006 and Age 0–4 prevalence data. Rate ratios are given for the log fits. Anova *p*-values are given for models comparing log gradients for: RR/SPMS and PPMS; 2006 and Age 0–4; and female and male for 2006. Cohort: Resident; Sample sizes: 765 RR/SPMS; 124 PPMS.

Latitude Slice (°)	MS Phenotype	Linear Regression Fit				Log Regression Fit						Anova <i>p</i> -value	
		Gradient	Std Err	R^2	<i>p</i> -value	Gradient	Std Err	R^2	<i>p</i> -value	Rate Ratio	Rate Ratio Range	RR/SPMS – PPMS	Male – Female
4.0	RR/SPMS	1.520	0.066	0.996	0.0019	0.144	0.026	0.939	0.0307	1.155	1.125–1.185	0.853	0.815
	PPMS	0.690	0.108	0.954	0.0235	0.153	0.039	0.886	0.0586	1.165	1.121–1.211		0.325
3.0	RR/SPMS	1.067	0.319	0.788	0.0442	0.090	0.038	0.659	0.0954	1.095	1.054–1.137	0.773	0.267
	PPMS	0.514	0.239	0.698	0.1646	0.108	0.039	0.793	0.1093	1.114	1.071–1.158		0.751
2.0	RR/SPMS	1.309	0.259	0.836	0.0039	0.108	0.030	0.724	0.0152	1.114	1.081–1.148	0.419	0.322
	PPMS	0.308	0.357	0.157	0.4374	0.046	0.075	0.085	0.5756	1.047	0.971–1.129		0.709

(a) Regression fits to the male 2006 data

Latitude Slice (°)	MS Phenotype	Linear Regression Fit				Log Regression Fit						Anova <i>p</i> -value	
		Gradient	Std Err	R^2	<i>p</i> -value	Gradient	Std Err	R^2	<i>p</i> -value	Rate Ratio	Rate Ratio Range	RR/SPMS – PPMS	2006 – Age 0–4
4.0	RR/SPMS	1.337	0.349	0.880	0.0618	0.097	0.029	0.849	0.0785	1.102	1.070–1.134	0.417	0.291
	PPMS	0.741	0.169	0.906	0.0482	0.126	0.015	0.974	0.0131	1.134	1.118–1.151		0.914
3.0	RR/SPMS	1.210	0.152	0.955	0.0041	0.095	0.011	0.963	0.0031	1.100	1.088–1.111	0.011	0.726
	PPMS	0.747	0.056	0.983	9.2e-4	0.146	0.009	0.989	5.1e-4	1.157	1.146–1.167		0.550
2.0	RR/SPMS	1.485	0.185	0.928	4.8e-4	0.120	0.015	0.931	4.4e-4	1.128	1.111–1.144	0.944	0.306
	PPMS	0.630	0.171	0.730	0.0144	0.122	0.028	0.795	0.0070	1.130	1.099–1.162		0.318

(b) Regression fits to the male Age 0–4 data

Table 5.11 Linear and log regression gradients—MS Phenotype (Male)

Gradients from linear and log regression fits to the male RR/SPMS and PPMS phenotypes, 2006 and Age 0–4 prevalence data. Rate ratios are given for the log fits. Anova *p*-values are given for models comparing log gradients for: RR/SPMS and PPMS; 2006 and Age 0–4; and female and male for Age 0–4. Cohort: Resident; Sample sizes: 190 RR/SPMS; 80 PPMS.

Although few of the observed differences in gradients given in Table 5.10 and Table 5.11 are statistically significant, based on the Anova models, it is interesting to look at how the rate ratios vary between factors. These comparisons point to combinations of factors where further analysis might produce significant results. Figure 5.20 compares rate ratios for the 2006 and Age 0–4 prevalence data. Despite variations in the rate ratios with latitude slice size, it can be seen that there does not appear to be any systematic difference in the rate ratios between the 2006 and the Age 0–4 prevalence data, for given combinations of gender and MS phenotype. Given the uncertainties involved, this is not inconsistent with the small drop in rate ratio observed in the

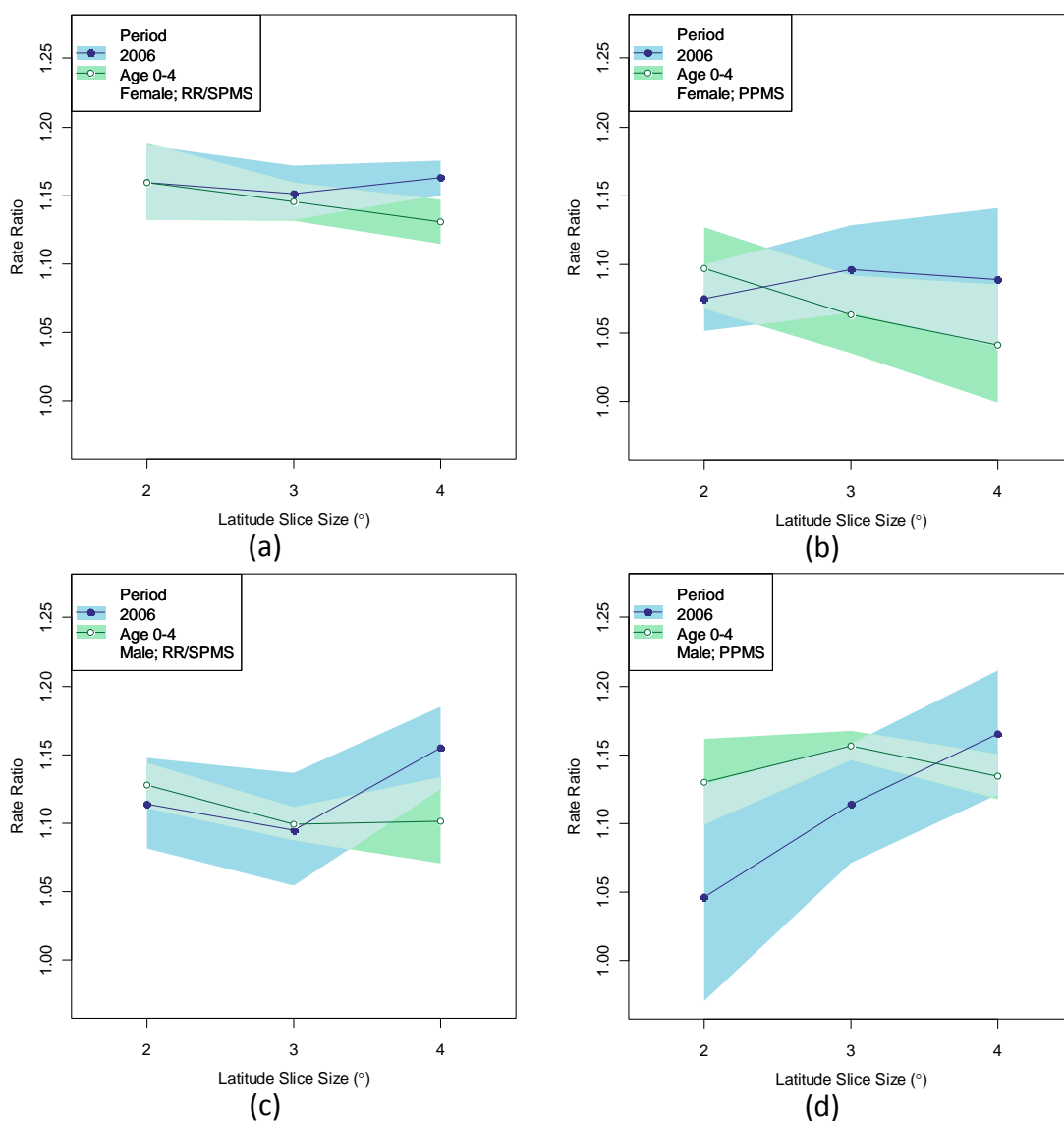


Figure 5.20 Rate Ratios—Life Period, Gender and MS Phenotype

MS prevalence rate ratios from log regression fits comparing 2006 prevalence and Age 0–4 prevalence. Shaded areas: 95% confidence bands for the rate ratios. **(a) (b)** Rate ratios for female cases for: **(a)** RR/SPMS; **(b)** PPMS. **(c) (d)** Rate ratios for male cases for: **(c)** RR/SPMS; **(d)** PPMS. (Standard: Euro; Latitude: PWC; Cohort: Resident.)

internal migration model considered in section 4.3.5 after applying the migration matrix. However, the slightly higher rate ratios seen for the 3° and 4° 2006 female data (compared with Age 0–4) suggest that there may be female specific risk factors which increase with latitude and act at a later age.

Figure 5.21 shows the rate ratios for the 2006 prevalence data. The only significant difference between rate ratios appears to be between RR/SPMS and PPMS phenotypes for female cases as shown in Figure 5.21(c). Although the difference appears to be consistent for all three latitude slice sizes, only the 2° comparison is statistically significant (Table 5.10(a)).

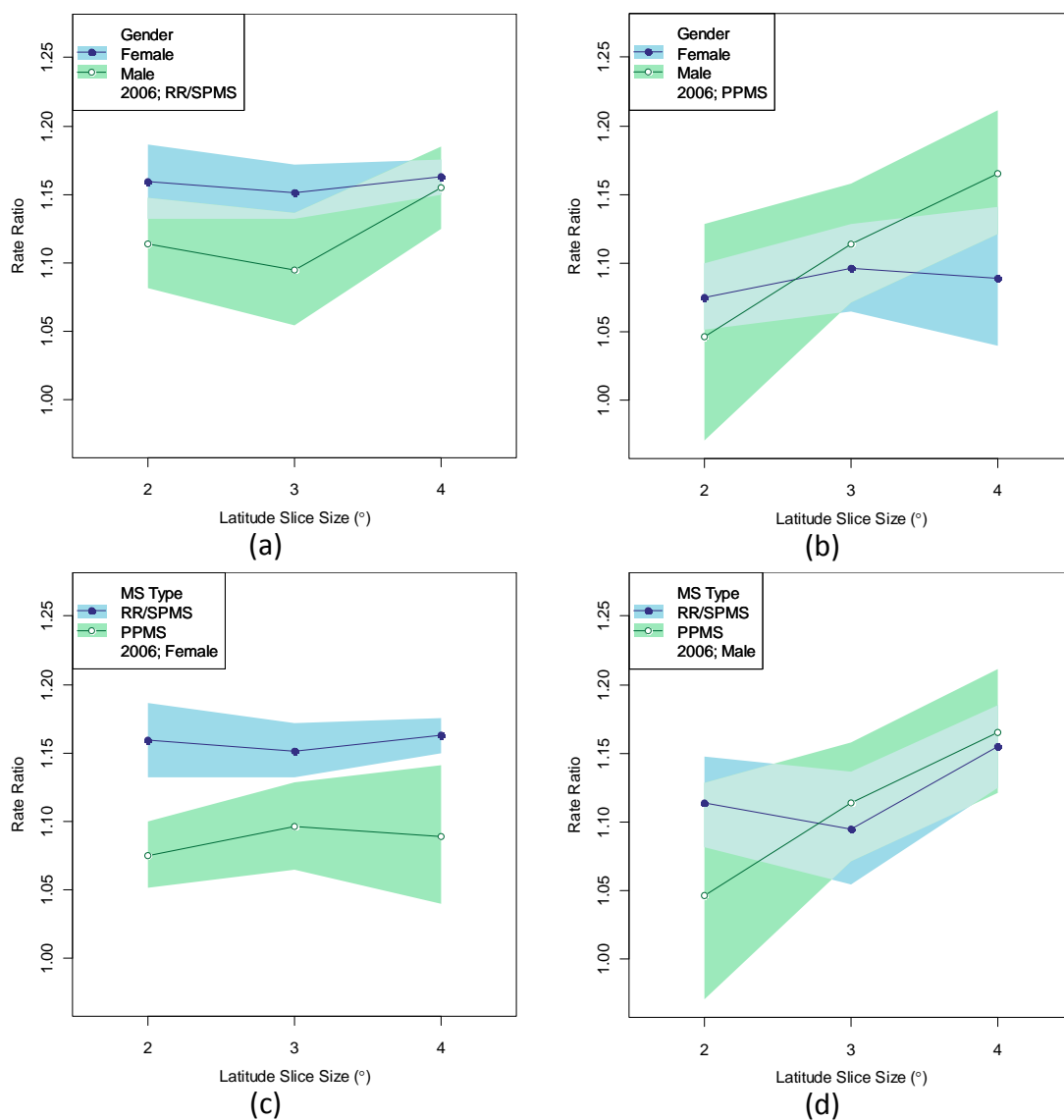


Figure 5.21 Rate Ratios—Gender and MS Phenotype (2006)

MS prevalence rate ratios from log regression fits for 2006 prevalence data. Shaded areas: 95% confidence bands for the rate ratios. **(a) (b)** Rate ratios for female and male cases for: **(a)** RR/SPMS; **(b)** PPMS. **(c) (d)** Rate ratios for RR/SPMS and PPMS for: **(c)** female cases; **(d)** male cases. (Standard: Euro; Latitude: PWC; Cohort: Resident.)

Figure 5.22 presents rate ratio data for Age 0–4 prevalence for the same combinations of other factors as for the 2006 data in Figure 5.21. In contrast to the 2006 data, there are now many differences in the rate ratios for the different combinations of gender and MS phenotype, and the choice of latitude slice size is also seen to be an important consideration. The 3° rate ratios are significantly different for each gender and MS phenotype comparison (Table 5.10(b) and Table 5.11(b)), however, none of the 2° or 4° comparisons are significant.

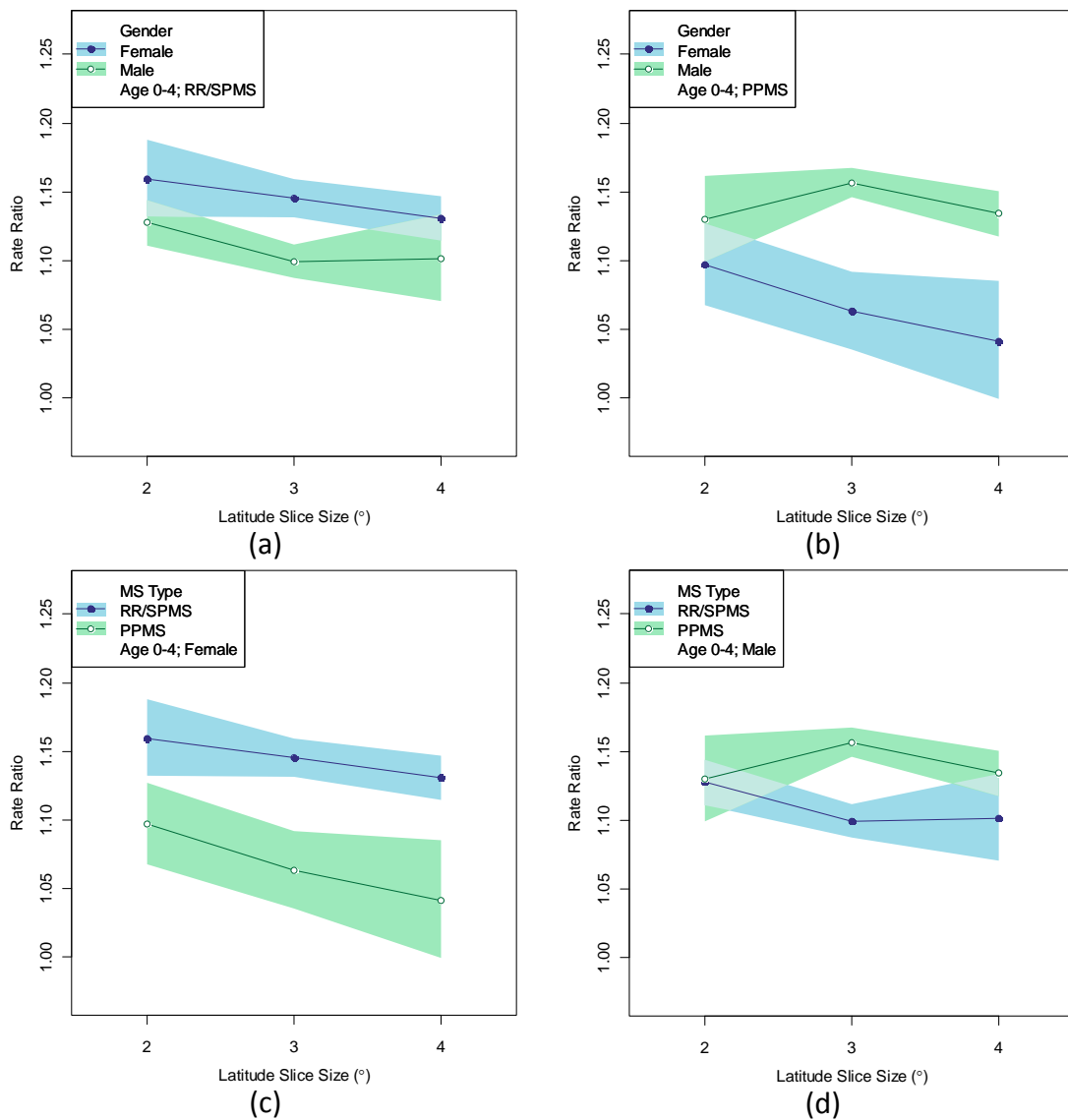


Figure 5.22 Rate Ratios—Gender and MS Phenotype (Age 0–4)

MS prevalence rate ratios from log regression fits for Age 0–4 prevalence data. Shaded areas: 95% confidence bands for the rate ratios. (a) (b) Rate ratios for female and male cases for: (a) RR/SPMS; (b) PPMS. (c) (d) Rate ratios for RR/SPMS and PPMS for: (c) female cases; (d) male cases. (Standard: Euro; Latitude: PWC; Cohort: Resident.)

It would be interesting to see whether different boundaries for the latitude slices affected the rate ratio results obtained either in terms of the figures obtained for the rate ratios or in terms of the significance of the results. The main difference between the 3° and 4° slices used in the analyses presented here is the location of Auckland City and the surrounding towns. For the 3° slices, Auckland falls near the centre of the 35°–38° slice, whereas for the 4° slices it falls relatively close to the boundary of the 33°–37° slice. Therefore, a initial comparison would be to analyse the data for 4° slices which included a 35°–39° slice. Subsequent comparisons might consider where slice boundaries fell in relation to other major cities such as Wellington and Christchurch, and alternative boundaries for the 3° and 2° slices. Note, however, that further investigations into the effect of the slice boundaries was not carried out in this study.

The results presented in Table 5.10 and Table 5.11 and illustrated in Figure 5.22 suggest that there are differences for the Age 0–4 rate ratios when considering different factor combinations, though with only the 3° slices giving differences which are statistically significant, no firm conclusions can be made. Bearing this in mind, however, it is useful to examine the differences which are observed by considering just the 3° rate ratios.

The comparisons shown in Figure 5.22(a) and Figure 5.22(b) suggest that, for the Age 0–4 data, the prevalence–latitude rate ratios are higher for males than for females for the PPMS phenotype, and higher for females than for males for the RR/SPMS phenotype. In addition, from the comparisons in Figure 5.22(c) and Figure 5.22(d), the prevalence–latitude rate ratios for females appear to be higher for the RR/SPMS phenotype than for the PPMS phenotype, whereas for males the rate ratios appear to be higher for the PPMS phenotype than for the RR/SPMS phenotype. For the 3° slice data these differences are statistically significant.

One question which comes to mind is this: how can the comparisons in Figure 5.22 show differences when few of the other comparisons do? To answer this take one of the comparisons as an example, without loss of generality. Consider the 3° slice data for the PPMS phenotype for which the rate ratios and Anova *p*-values are given in the fourth lines of each of Table 5.10(a) and (b) and Table 5.11(a) and (b). The data is also repeated in Table 5.12 for clarity. Each successive entry in Table 5.12 has one change in the combination of factors, and although the rate ratios are seen to gradually

Gender	Period	Rate Ratio	Rate Ratio Range	Anova <i>p</i> -value	
Female	Age 0–4	1.063	1.036–1.092	} 0.509	\ 0.023
Female	2006	1.096	1.065–1.128		
Male	2006	1.114	1.071–1.158	} 0.751	
Male	Age 0–4	1.157	1.146–1.167	} 0.550	

Table 5.12 Rate Ratio Comparisons for PPMS Phenotype and 3° latitude slices
Rate ratio data and Anova comparisons from Table 5.10 and Table 5.11 for the PPMS phenotype and 3° PWC latitude slices. The first column of Anova *p*-values compare log regression fits for successive table entries; the final value is for a comparison between the first and last entries in the table. (Standard: Euro; Cohort: Resident.)

increase as the factors change, none of these individual changes produce significant differences in rate ratios. However, the difference between the first and last combinations, the fits which give the lowest and highest rate ratios, is significant. It can also be seen that, although the successive rate ratio changes combine to produce a significant difference, the cumulative change in factors only amounts to a change in gender from female to male—the first change from Age 0–4 prevalence to 2006 prevalence is reversed later. Thus, the differences in rate ratios for Age 0–4 between female and male cases for the 3° latitude slices and PPMS phenotype can be seen to be significant despite there being no significant differences for the rate ratios of other combinations of factors.

The results presented in this section show that the prevalence rates of MS, when taking gender and MS phenotype into consideration, appear to depend differently on latitude when looking at residence location at Age 0–4 compared with results based on residence location on census day 2006. Therefore, although migration is not strictly a factor (as discussed in Chapter 3), it is useful to look at migration next. First, the Migrated cohort of cases excluded from the analyses of this section will be analysed further, and then internal migration patterns of the cases included will be examined.

5.3.2 Migration—External

The Resident cohort of cases analysed in the previous sections only included cases who had been continually resident in New Zealand from birth to onset of MS, in order to

control the effect of external migration. In addition, questionnaire data returned by cases who had resided outside of New Zealand often only included a country of residence rather than specific locations, so any analyses of latitude of residence at different life periods would be more problematic. However, given that the results presented in the previous sections indicate that latitude (as a surrogate) has different effects on prevalence of MS for the different gender–MS phenotype combinations at different ages, possibly due to internal migration effects (discussed further in the next section) it is therefore useful to look at the effects of gender and MS phenotype on the 2006 prevalence–latitude gradients of the Migrated cohort of cases which was excluded from that analysis. This will clarify how much of an effect migration to and from New Zealand has on the prevalence of MS. Note that only data aggregated to 3° latitude slices and age-standardised to the European standard will be analysed, and only the log regression fits to the data will be presented.

Figure 5.23 shows the prevalence of MS with latitude for the Migrated cohort for the four gender–MS phenotype combinations, with the log regression data and Anova *p*-values for log gradient comparisons given in Table 5.13(a) and (c). Table 5.13(b) gives the log regression data for the Resident cohort (taken from Table 5.10(a) and Table 5.11(a), repeated for convenience); Table 5.13(c) also gives Anova *p*-values for comparisons between the log gradients for the Resident and Migrated cohorts. The log regression fits shown in Figure 5.23 for the Migrated cohort are equivalent to those in Figure 5.16(c) and (d) and Figure 5.18(c) and (d) for the Resident cohort. Despite lower sample numbers for the Migrated cohort compared to the Resident cohort, the log fits for all but the female PPMS data have lower uncertainties and narrower confidence bands. For the female RR/SPMS cases, the rate of change of prevalence with latitude is significantly lower for the Migrated cohort than for the Resident cohort, whereas for the male PPMS cases, the rate for the Migrated cohort is significantly higher than that for the Resident cohort. This implies that there are competing influences acting separately by gender and by MS phenotype. Some factor or factors associated with migration reduce the effect of latitude in female RR/SPMS cases; other factors increase the effect of latitude in male PPMS cases. For female PPMS and male RR/SPMS cases these influences broadly cancel out. Further stratification of the data by the period spent outside of New Zealand, and residence location, would help to clarify factors involved since it is likely that the prevalence–latitude relationships observed are

themselves combinations due to other factors. However, the sample numbers which would be involved for this cohort of cases would make such an analysis unrealistic, though an approach similar to that used for month of birth data (see section 5.3.4), where a simple north–south divide is utilised, might be feasible. This analysis option has not been pursued here.

Within the Migrated cohort, the log gradient for the male PPMS cases is significantly different from that of both the female PPMS cases and the male RR/SPMS cases. This can also be seen from the rate ratios: the female PPMS rate ratio is only 1.035 and the

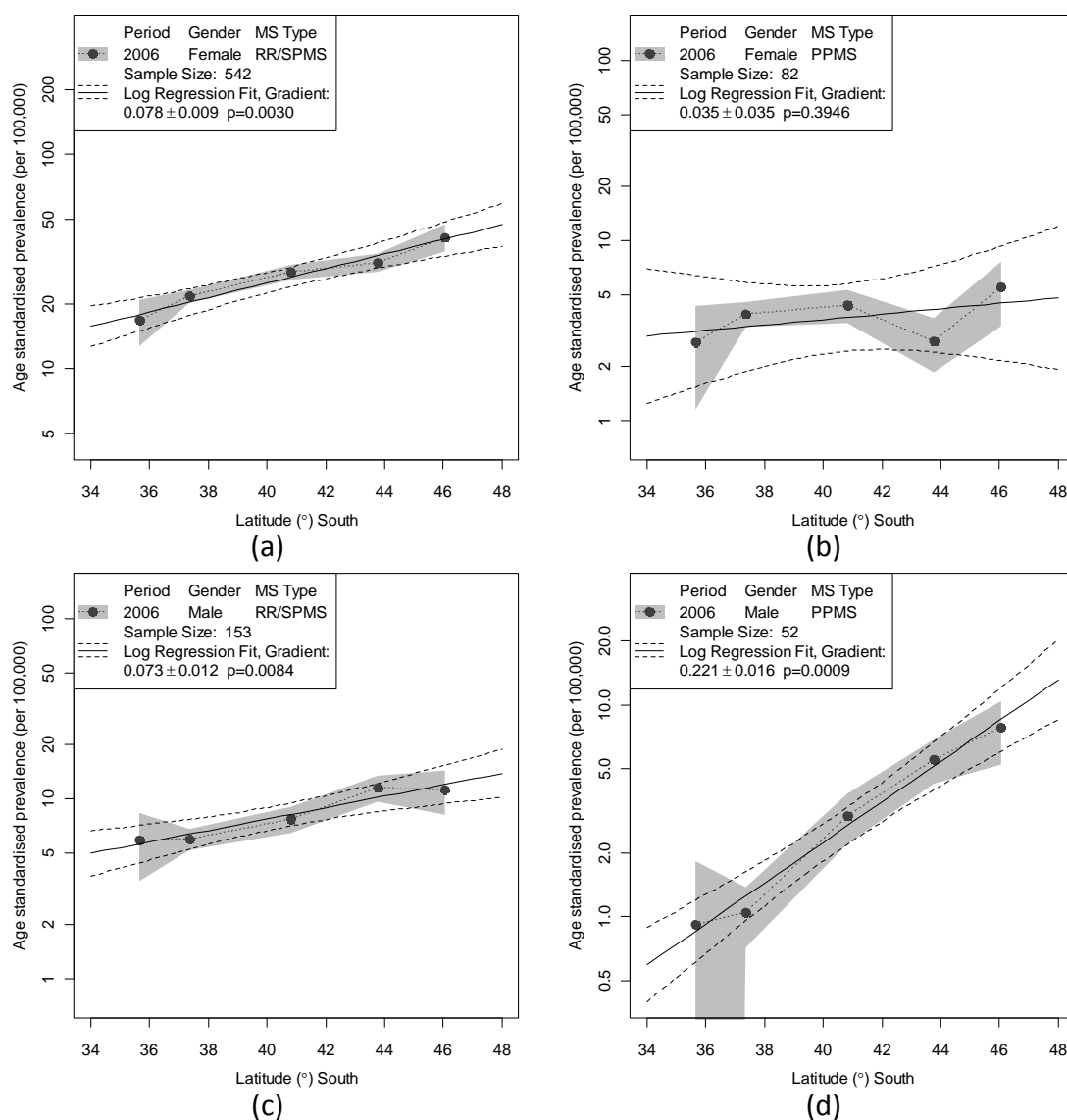


Figure 5.23 Log regression fits—2006 Prevalence for Migrated cohort

Log regression fits for the 2006 prevalence data for the Migrated cohort, by gender and MS phenotype. Shaded areas: standard error uncertainties in the data; solid lines: regression fits; dashed lines: 95% confidence bands for the regression fits. Gender: **(a) (b)** female, **(c) (d)** male; MS phenotypes: **(a) (c)** RR/SPMS **(b) (d)** PPMS. (Standard: Euro; Latitude: PWC; Latitude slices: 3°.)

Gender	MS Phenotype	Log Fit Gradient	Std Err	R^2	p -value	Rate Ratio	Rate Ratio Range
Female	RR/SPMS	0.078	0.009	0.964	0.0030	1.081	1.072–1.091
	PPMS	0.035	0.035	0.247	0.3946	1.035	1.000–1.072
Male	RR/SPMS	0.073	0.012	0.928	0.0084	1.076	1.063–1.089
	PPMS	0.221	0.016	0.984	0.0009	1.247	1.227–1.268

(a) Data: 2006, Migrated

Gender	MS Phenotype	Log Fit Gradient	Std Err	R^2	p -value	Rate Ratio	Rate Ratio Range
Female	RR/SPMS	0.141	0.017	0.957	0.0038	1.152	1.132–1.172
	PPMS	0.092	0.029	0.771	0.0502	1.096	1.065–1.128
Male	RR/SPMS	0.090	0.038	0.659	0.0954	1.095	1.054–1.137
	PPMS	0.108	0.039	0.793	0.1093	1.114	1.071–1.158

(b) Data: 2006, Resident

	Female	Male	Resident	Migrated
	Comparison: Resident–Migrated		Comparison: Female–Male	
RR/SPMS	0.017	0.675	0.267	0.753
PPMS	0.255	0.029	0.751	0.003
	Comparison: RR/SPMS–PPMS			
Resident	0.193	0.773		
Migrated	0.276	3.3e-4		

(c) Anova p -values for the given comparisons

Table 5.13 Log regression gradients—2006 Prevalence for Migrated cohort

(a) Gradients and rate ratios from log regression fits to the 2006 prevalence data for the Migrated cohort. (b) Equivalent data for the Resident cohort taken from Table 5.10(a) and Table 5.11(a). (c) Anova p -values for models comparing log gradients for the given combinations of gender and MS phenotypes. Resident and Migrated cohorts.

male RR/SPMS rate ratio is 1.076, giving percentage increases in prevalence of MS per degree of latitude of 3.5% and 7.6% respectively. In contrast, the male PPMS rate ratio is 1.247, almost equivalent to a 25% increase in prevalence per degree of latitude. This high rate ratio amounts to a tenfold increase in the prevalence of MS in the far south of New Zealand compared to the far north, for this cohort.

Within the Resident cohort, only the comparison between female RR/SPMS and female PPMS log gradients for 2° latitude slices gave a significant difference, with suggestions that the rate ratios obtained depended upon the latitude slice size used. Here, for the Migrated cohort, the data has only been analysed using 3° slices. A more thorough analysis looking at the effect of latitude slice size would also be prudent, though this has not been carried out at this time.

Migration to and from New Zealand does affect the prevalence of MS within New Zealand and the effect varies with gender and MS phenotype. The next section will examine whether migration within New Zealand is also likely to have an effect.

5.3.3 Migration—Internal

The matrix analysis outlined in section 4.3.5 suggests that natural migration processes within a population structure such as that present in New Zealand, should reduce the south-north rate ratio seen for a given cohort of MS cases over time. That is, the generally higher populations towards the north of the country attract more internal migrants from the south than are balanced by the reverse migration. The distribution of MS cases would also follow this trend so that more of the cases in the south would tend to migrate north, thus reducing the rate ratio. However, although the shift in population which is observed is the result of migration in both directions, in this analysis, the majority of the population does not migrate any significant distance. Therefore, any migration effects which are present in the data from the Resident cohort of cases studied here may be overshadowed by the lack of any effect in the rest of the data. This is consistent with the lack of any systematic difference in the rate ratios when comparing the rate ratios for the census day 2006 and Age 0–4 prevalence data, as shown in the previous section in Figure 5.20, though as mentioned in the discussion of Figure 5.20, there is a suggestion of female specific factors acting at later ages. However, there is also the question of whether or not prevalence at census day 2006 can be taken as indicative of prevalence prior to onset of MS: the following analysis of the migration patterns of the Resident cohort of MS cases suggests that it cannot due to migration after onset of MS.

The main period of concern here is that between birth (Age 0–4) and onset of MS. Any effects which migration may have on an individual case are likely to be greater for a longer net migration over this period compared to a shorter migration. (Net migration is considered to be the difference between residence locations at birth and at onset of MS, regardless of the route taken in between.) To measure longer net migration, an arbitrary threshold of 2° of latitude was applied and the residence location data was analysed for net migration greater than this threshold, that is, for MS cases who had migrated more than 2° latitude. Table 5.14 shows the number of MS cases where net migration was more than 2° latitude, north or south. Data is given separately for (a)

the pre-onset period from birth to onset of MS, (b) the post-onset period from onset of MS to census day 2006 and (c) the whole period from birth to census day 2006. For the pre-onset period, less than 20% of MS cases migrated more than 2° and it is noticeable that a higher proportion of male cases than female cases migrated this distance. Also, more cases migrated north rather than south, however, for females the ratio is 1.45:1 compared with only 1.10:1 for males. For the period after onset of MS and before census day 2006, there is little difference in the proportions of females and males migrating north more than 2°, with a higher proportion of males than females heading south. It is interesting to note that the numbers migrating south prior to onset of MS are almost the same as the numbers for the whole period up to census day 2006,

	All Cases	> 2° north	> 2° south	% north	% south	N:S
All	1166	119	89	10.2	7.6	1.34:1
Female	894	87	60	9.7	6.7	1.45:1
Male	272	32	29	11.8	10.7	1.10:1
% Female	76.7	73.1	67.4			
% Male	23.3	26.9	32.6			
F:M	3.29:1	2.72:1	2.07:1			
(a)						
	All Cases	> 2° north	> 2° south	% north	% south	N:S
All	1166	88	40	7.5	3.4	2.20:1
Female	894	68	29	7.6	3.2	2.34:1
Male	272	20	11	7.4	4.0	1.82:1
% Female	76.7	77.3	72.5			
% Male	23.3	22.7	27.5			
F:M	3.29:1	3.40:1	2.64:1			
(b)						
	All Cases	> 2° north	> 2° south	% north	% south	N:S
All	1166	169	89	14.5	7.6	1.90:1
Female	894	125	62	14.0	6.9	2.02:1
Male	272	44	27	16.2	9.9	1.63:1
% Female	76.7	74.0	69.7			
% Male	23.3	26.0	30.3			
F:M	3.29:1	2.84:1	2.30:1			
(c)						

Table 5.14 Net Migration—Gender

The number of cases of MS who migrated more than 2° latitude north or south from: **(a)** birth to onset of MS; **(b)** onset of MS to census day 2006; **(c)** birth to census day 2006. F:M : female:male; N:S : north:south. All cases were resident in New Zealand from birth to onset of MS; migration is net migration for the period concerned.

which suggests that many who had migrated south prior to onset of MS migrated again more than 2° after onset, though whether this subsequent migration was north or south is not clear from this particular analysis. For those who migrated north more than 2° prior to onset of MS, fewer migrated again more than 2° after onset.

Migration can also be categorised by MS phenotype as well as gender, though the small numbers of cases involved increases the uncertainties in any comparisons which are made. (Note that no attempt has been made here to establish the uncertainties in the data.) However, several further comparisons are worth noting for the data for the birth to onset of MS period, as given in Table 5.15. For male cases, the north:south ratio for RR/SPMS is 0.71 (17:24) and for PPMS it is 3.00 (15:5). Although the numbers are small, this means that for migration of more than 2° latitude, over 40% more male RR/SPMS cases migrated south rather than north, whereas three times as many PPMS cases migrated north rather than south. In contrast, an equal number of female PPMS cases migrated north as migrated south, and 56% more RR/SPMS migrated north rather than south.

RR/SPMS	All Cases	> 2° north	> 2° south	% north	% south	N:S
All	955	92	72	9.6	7.5	1.28:1
Female	765	75	48	9.8	6.3	1.56:1
Male	190	17	24	8.9	12.6	0.71:1
% Female	76.7	81.5	66.7			
% Male	23.3	18.5	33.3			
F:M	4.03:1	4.41:1	2.00:1			
(a)						
PPMS	All Cases	> 2° north	> 2° south	% north	% south	N:S
All	204	27	17	13.2	8.3	1.59:1
Female	124	12	12	9.7	9.7	1.00:1
Male	80	15	5	18.8	6.3	3.00:1
% Female	76.7	44.4	70.6			
% Male	23.3	55.6	29.4			
F:M	1.55:1	0.80:1	2.40:1			
(b)						

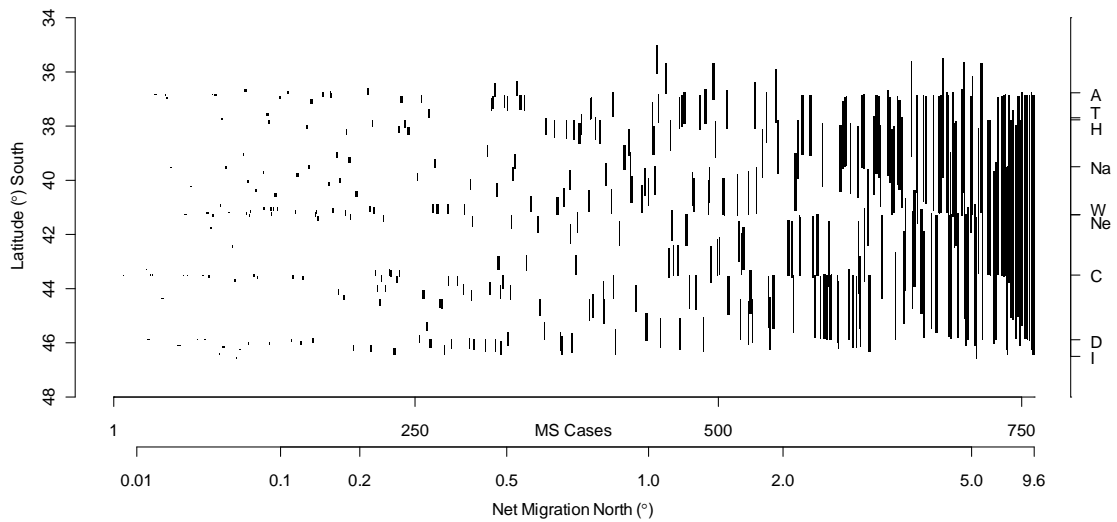
Table 5.15 Net Migration—Gender and MS Phenotype

The number of cases of MS who migrated more than 2° latitude north or south from birth to onset of MS: **(a)** RR/SPMS; **(b)** PPMS. F:M : female:male; N:S : north:south. All cases were resident in New Zealand from birth to onset of MS; migration is net migration for the period concerned.

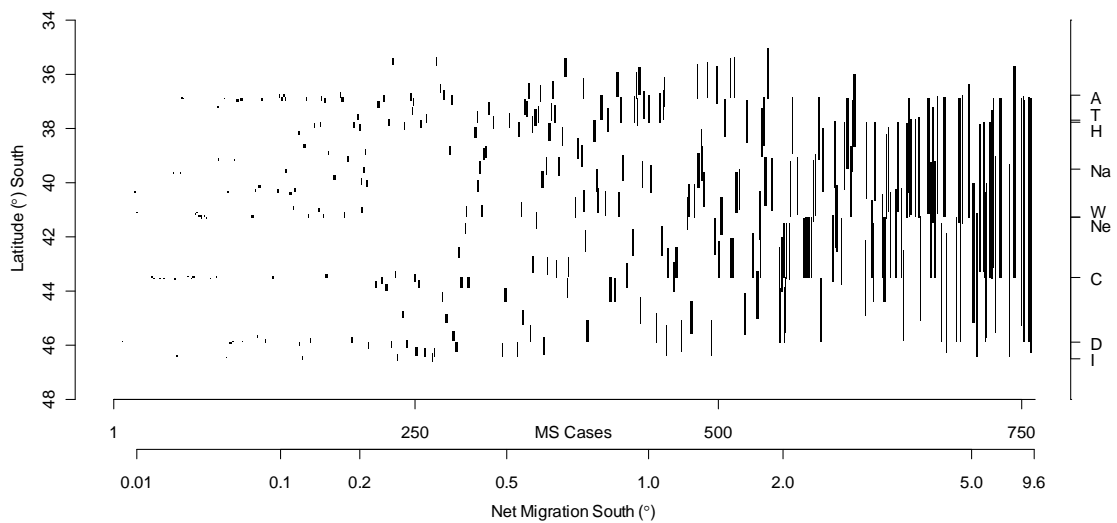
It should also be noted that the main cities of New Zealand, with the exception of Invercargill to the south and Auckland to the north, are all roughly equally spaced along the length of New Zealand, and at just above or below 2° latitude spacing. (As will be illustrated below in Figure 5.24 and Figure 5.25, Hamilton and Dunedin are 8.1° latitude apart, with Christchurch, Nelson/Wellington and Napier located in between.) Whether the 2° latitude threshold used above is suitable in view of this, will not be discussed here.

The complexities of the variations in the numbers of cases who migrated more than 2° for the different life periods of concern here, suggest that an interpretation of the observed prevalence–latitude rate ratios in terms of migration is not straightforward, especially given that rate ratios are not available from this study for the life period immediately prior to onset of MS. The extent to which the variations in migration are correlated with prevalence cannot therefore be investigated. It is possible, however, to establish other trends from the migration data which may assist future studies.

The migration patterns of the MS cases, without regard to gender or MS phenotype, are illustrated in Figure 5.24 for the pre-onset period, and Figure 5.25 for the post-onset period. (Note that although the same horizontal scale is used, with cases ordered by net migration, there is no correlation between Figure 5.24 and Figure 5.25 in terms of the position of any individual case.) Lines are marked between the latitudes of residence locations at birth and at onset of MS, or onset of MS and at prevalence 2006, respectively. The horizontal axes have the same scale in both (a) and (b) in both figures and extends to cover all cases, ordered by net migration either north or south. For clarity, however, cases with a net migration north are shown in (a) and cases with a net migration south are shown in (b). For reference, the locations of major cities are shown on the right hand axes. For pre-onset (Figure 5.24), the 2° threshold used to look at migration patterns earlier in this section captures the 26% of MS cases with the greatest net migration either north or south (excluding cases with no net migration); for post-onset (Figure 5.25) the 2° threshold captures 17%. So for most cases, the magnitude of net migration is less post-onset than for the pre-onset period. However, it can also be seen that almost as many MS cases migrated after onset of MS (738) as did prior to onset (760).



(a)



(b)

Figure 5.24 Net migration within New Zealand—Pre Onset

Net migration of each MS case is shown sequentially from left to right, ordered by net latitude migrated. A case is either in **(a)**, with a space in **(b)**, or in **(b)**, with a space in **(a)**. Vertical bars span the latitudes migrated from birth to onset of MS.

Net migration: **(a)** north; **(b)** south. Indicated on the right are approximate locations of: A - Auckland; C - Christchurch; D - Dunedin; H - Hamilton; I - Invercargill; Na - Napier; Ne - Nelson; T - Tauranga; W - Wellington.

It is noticeable in both figures that where cases have a larger net migration, either the location at birth or the location at onset of MS is likely to be in the vicinity of one of New Zealand's main cities, especially Auckland. Shorter migration spans also correlate with cities to some extent, however, there is a much more even distribution of latitudes involved in migration spans of between 0.5° and 1°. Shorter migrations again show a correlation with the locations of cities, suggesting that many may migrate around the area of a city rather than migrate to another city.

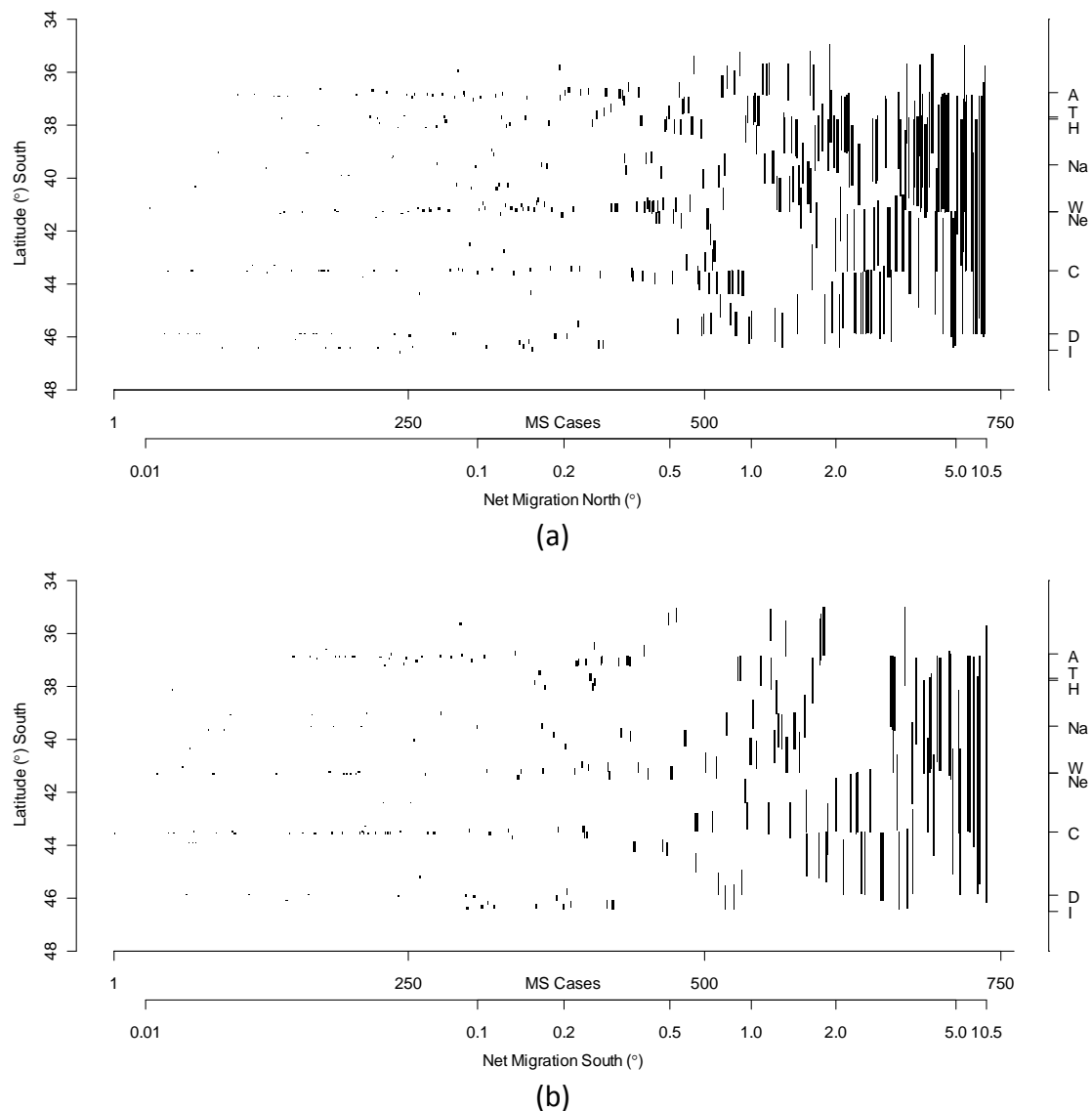


Figure 5.25 Net migration within New Zealand—Post Onset

Net migration of each MS case is shown sequentially from left to right, ordered by net latitude migrated. A case is either in **(a)**, with a space in **(b)**, or in **(b)**, with a space in **(a)**. Vertical bars span the latitudes migrated from onset of MS to census day 2006. Net migration: **(a)** north; **(b)** south. Indicated on the right are approximate locations of: A - Auckland; C - Christchurch; D - Dunedin; H - Hamilton; I - Invercargill; Na - Napier; Ne - Nelson; T - Tauranga; W - Wellington.

The main question being asked here is whether the prevalence data for census day 2006 can be used in lieu of prevalence data for the period prior to onset of MS. The data presented in Table 5.14 and Figure 5.25 suggests that net migration after onset of MS and prior to census day 2006 would be sufficient to confound any effects of net migration prior to onset of MS, and consequently, that the prevalence of MS on census day 2006 cannot be used to look at migration effects prior to onset of MS. This implies that further analyses to determine prevalence at onset of MS is required before an in depth assessment of the effects of migration can be made.

5.3.4 Month of Birth

As discussed in Chapter 2, there is evidence that there is a 'month of birth' effect when looking at the prevalence of MS. However, there is also a 'month of birth' effect in the general population. It is necessary, therefore, to establish whether the effect within the cases of MS follow the same trend as the general population, or whether the effect cannot be attributed to this and, as such, must be attributed to some other factors.

The mean year of birth for the cohort of MS cases studied here has been established as being around 1954 but with a considerable spread of years as shown in Figure 4.1(a).

The underlying general population trend for a single year would therefore not be appropriate (even if available), instead it is necessary to consider an average trend for a number of years.

Data is available from StatsNZ (StatsNZ n.d.) which gives the numbers born in each month for the whole of New Zealand, as shown in Figure 5.26(a). However, this data only goes back to 1980 and, although there is a monthly pattern within each year, there is also significant variation from year to year. In addition, the monthly pattern appears to be much less regular in the years after around 1999. Although any attempt to extrapolate this birth data back to 1954, or any other year, would not be feasible, an annual trend can be established by normalising each year's data by the total number of births for that year. After allowing for the number of days in each month, the 5-year means for each month are as shown in Figure 5.26(b), here expressed as percentage differences from the annual monthly mean. Only years up to 1999 are considered, and the 20-year monthly means for 1980–1999 are also shown. These 20-year means are then used in further analyses—assuming that the population within each census area, and hence within each latitude slice, follows this trend, a population by month of birth can be estimated. (Note that no attempt is made to estimate the uncertainties involved in these monthly population estimates.)

Section 5.3.1 demonstrated that some latitude effects were only apparent after the data was stratified by both gender and MS phenotype. Unfortunately, this degree of stratification greatly reduces the number of cases per category. If further stratification is to be considered by month of birth then the numbers involved are unlikely to produce viable results. Since the effect of latitude within the data presented here is well established, one option to improve case numbers per category is to reduce the

number of latitude slices being considered. If just two slices are considered (the minimum without discarding all latitude information) this amounts to splitting the country across a nominal 'middle' to give 'north' and 'south' areas. Although detailed analyses by latitude would not be possible, any latitude effect should still be apparent: variations which appear consistently in both the north and the south data are likely to be latitude–month of birth effects; variations which differ from north to south would be inconclusive—they could either be due to natural uncertainties in the data, or a month of birth latitude effect, or a combination of both. Since the New Zealand area of interest extends approximately from 34° to 48° south, a division at 41° south was used

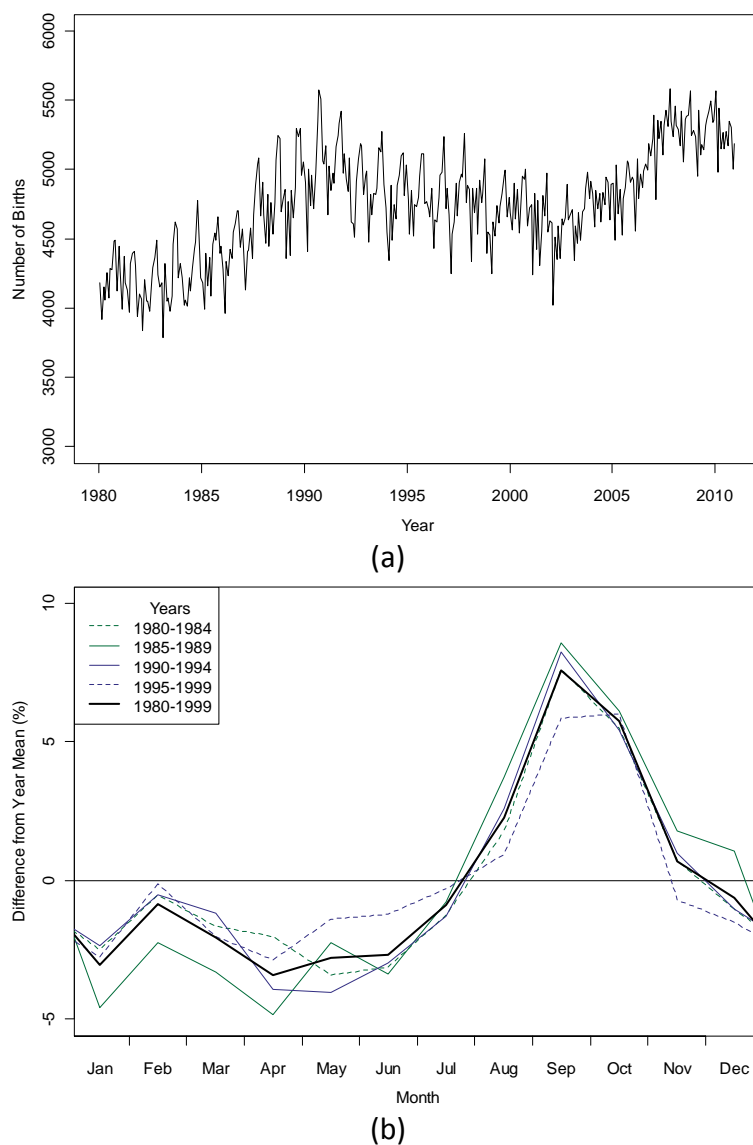


Figure 5.26 Number of births in New Zealand by month of birth

Births in New Zealand per month of birth. **(a)** The number for each month from Jan 1980 to Dec 2010 (StatsNZ n.d.). **(b)** Five year means for years up to 1999 and the 20 year mean for 1980–1999 (see text for details). Means for each month are expressed as differences from the annual monthly means.

for the following analyses. Data for New Zealand as a whole, without differentiating north and south, was also analysed for comparison. In addition, for the following analyses, only prevalence at Age 0–4 for the Resident cohort is considered and crude prevalence is reported rather than age-standardised data.

Figure 5.27 shows crude prevalence by month of birth, taking into account the 20-year population month of birth trend as given in Figure 5.26(b). (The different numbers of days in each month are taken into account since the numbers cancel out between the prevalence data and the birth rate data.). Data is shown separately by gender and MS phenotype. Note that it is the variations in numbers from month to month within each category (gender–phenotype combination) which are of interest here, not changes in

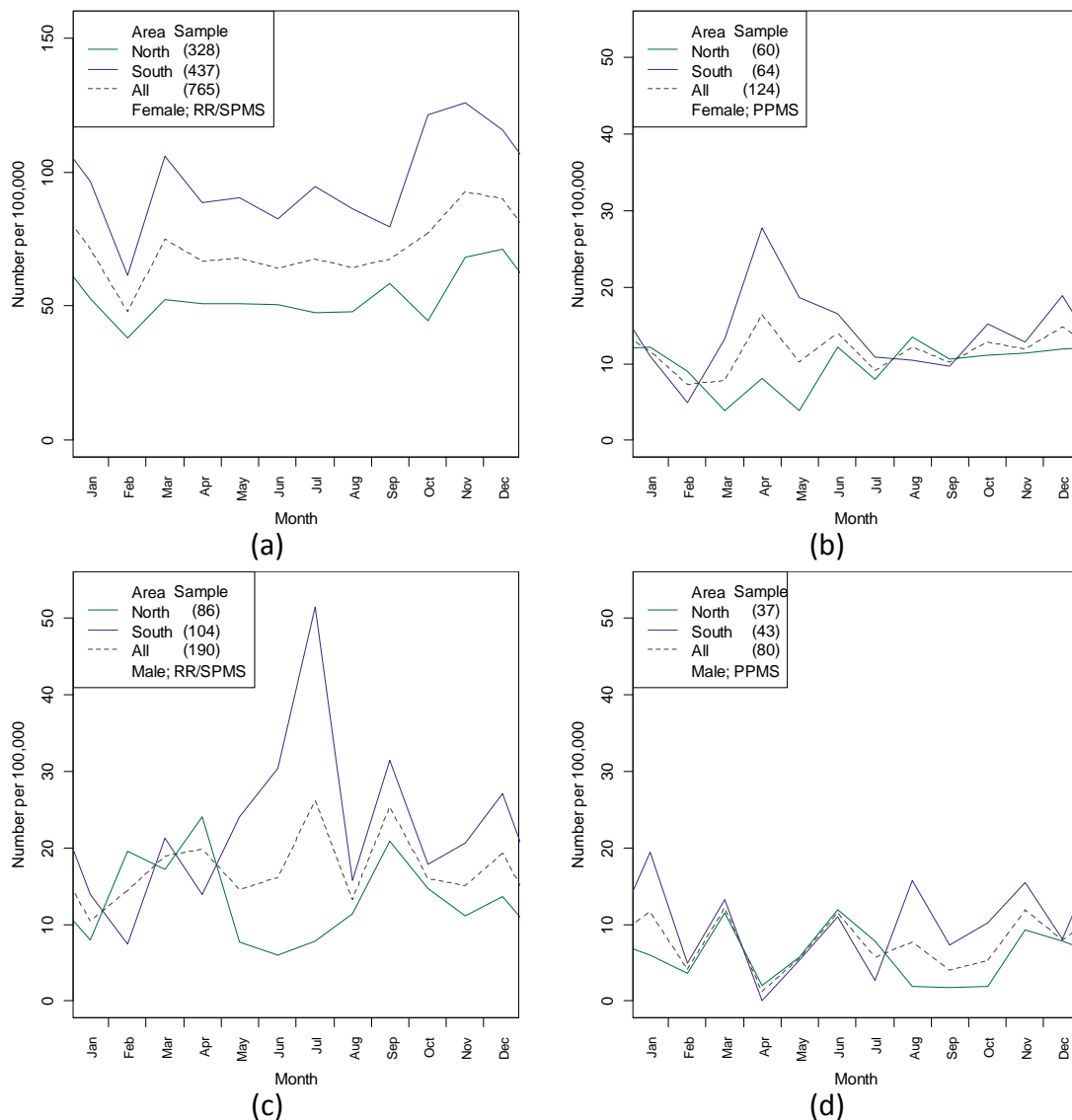


Figure 5.27 Crude prevalence by Month of Birth
Crude MS prevalence rates at Age 0–4 by month of birth. North: north of 41°S; south: south of 41°S. **(a)** Female RR/SPMS cases; **(b)** female PPMS cases; **(c)** male RR/SPMS cases; **(d)** male PPMS cases. (Cohort: Resident.)

scale between categories, though it is very noticeable that the patterns of variations in the data differ between the four categories. Note also, that if these annual variations are linked with variations in UV exposure and to measures such as vitamin D status (as discussed in Chapter 2) then any lag between low UV exposure and low vitamin D blood serum levels during winter and spring must be taken into consideration. This could place periods of interest for low vitamin D status up to several months later than stated below.

Each category will now be discussed in turn, starting with the female RR/SPMS data which has the largest sample size and, hence, which should give the most significant results.

The female RR/SPMS data (Figure 5.27(a)) is distinctive for two reasons. First, the north and the south data do not overlap, which is consistent with the high rate ratios shown in Figure 5.22(a). Second, there appears to be a sharp drop in February and a broad peak from November to December in the north and from October to December in the south. January appears to be a natural progression from the November peak to the February drop. However, March appears to show a slight peak compared to the rather steady rates from April through to August. All of these month effects are stronger for the south than for the north. If the peak in numbers around November is a real feature then it is interesting to note that this corresponds to a second trimester of June to August, which is mid winter. In addition, the drop in birth numbers of MS cases in February corresponds to births just after mid summer, which suggests that there is a beneficial effect in the month or two prior to birth.

The sample numbers for the female RR/SPMS cases are reasonable for this month of birth analysis, with an average of around 32 births per month, however, those for male RR/SPMS cases (Figure 5.27(c)) are just under 8 per month, so natural variations and uncertainties in the data are likely to be more pronounced. Despite this, there is seen to be a strong peak in the prevalence rates for July in the south, with a corresponding broad drop in the rates around June in the north. Whether this pattern could be due to random variations, or whether it is likely to be significant was not investigated. It is interesting that for the average rates for the whole of New Zealand these two features average out sufficiently so that the variations in the average could appear to be due to natural variations and uncertainties in the data. However, a comparison is made,

below, with data from Australia which suggests that the broad drop in June in the north may be a real effect. This would correspond to a beneficial effect of mid summer, occurring in the second trimester. The peak in the south at the same time would correspond to birth during mid winter. This suggests that there may be competing effects which act at different times during pregnancy and which vary depending upon latitude.

The prevalence rates for the PPMS cases are shown in Figure 5.27(b) for female cases and Figure 5.27(d) for male cases, though now the sample numbers average 5.2 and 3.3 per month, respectively. Although there is a suggestion of a peak around April for the south female PPMS cases, with a corresponding low from March to May in the north, there appears to be no other evidence of any month of birth effect for this category. For the male PPMS cases, there appears to be a correlation between rates for the north and for the south for February to June. However, the lows in April correspond to just one case for the north and no cases for the south, and the peaks in March and June for north and south each have just four to six cases. These small numbers suggest that the apparent correlation could easily be due to chance, though again, this was not tested here. The overall annual trend in the data for the south suggests a low around April to May, however this is not conclusive. It would appear, therefore, that there are no notable month of birth effects within these cases.

The only feature which appears to be consistent across genders and MS phenotypes is a low number of February births in the South. This could be due to random variations for the low sample number for male PPMS cases, though for the other three gender–MS phenotype combinations sample numbers are not so much of an issue.

Although the month of birth data was not considered here without stratifying by MS phenotype, some comparison can be made with data from southern Australia. For latitudes from 27.5°S to 42.9°S, Staples et al (2010) found an increase in prevalence of all MS for the two month Nov–Dec period, and a drop in the May–June period. The female RR/SPMS data presented here (Figure 5.22(a)) shows a peak in November which could correspond to the peak in the Australian data, but no corresponding drop in May–June. However, the northern male RR/SPMS data (Figure 5.22(c)) does show a broad drop from May to July, as mentioned above. This suggests that there may be real differences in month of birth effects which depend on MS phenotype.

As discussed in Chapter 2, it is likely that any prevalence–month of birth effects vary by gender and MS phenotype. Unfortunately, sample numbers only allow reliable rates to be estimated for female RR/SPMS cases. For all other cases, uncertainties in the data may be greater than any effects which might otherwise be present. As well as looking further at the uncertainties in the data in order to ascertain which of the observed effects are likely to be significant, smoothing techniques such as a simple running average or a more complex locally weighted smoothing algorithm could also be utilised to help in reducing the variations in the data. However, care must also be taken so that real features are not smoothed out too far and lost in the process. Similarly, ignoring the effect of MS phenotype and looking at all phenotypes together may increase the sample numbers and reduce uncertainties, but it also increases the risk of averaging out effects which might be present but which differ by MS phenotype.

5.4 Summary of Latitude Results

It has been established that there is a correlation between the prevalence of MS within New Zealand and the latitude of residence, with higher (more southerly) latitudes having the highest prevalence rates. This correlation can be expressed in terms of the rate ratio obtained from log regression fits to the prevalence–latitude data. For the prevalence on census day 7 March 2006, the rate ratio obtained depends on whether or not cases who have migrated to or from New Zealand are included. The increase in prevalence of MS per degree of latitude south is around 12% for all cases, 8% for cases who have resided outside of New Zealand at some point prior to onset of MS (the Migrated cohort), and 14% for cases who have resided in New Zealand from birth to onset of MS (the Resident cohort).

The rate ratios obtained are seen to vary, depending on a number of other factors, in particular, gender and MS phenotype. For the Migrated cohort, the increase in prevalence of MS per degree of latitude south is around 3.5% for female PPMS cases, 8% for female and male RR/SPMS cases, and 25% for male PPMS cases. For the Resident cohort, the equivalent rates vary from 10% to 15%. Further factors for the Resident cohort (not analysed for the Migrated cohort) include the period of life under consideration, and the areal interpolation method used. The choice of standard population for the age standardisation process used in determining the prevalence rates has little effect on the rate ratios reported here.

Migration to and from New Zealand is seen to be a major factor in determining the rate at which prevalence of MS changes with latitude within New Zealand. In addition, migration appears to influence the prevalence of MS differently according to gender and MS phenotype.

Internal migration is also shown to be an important confounding factor for establishing a link between latitude and prevalence of MS. Significant differences in rate ratios for different factors can be seen for residence location at Age 0–4; these differences are mostly absent when looking at residence location on census day 2006. An examination of the patterns of migration concludes that the prevalence rates for census day 2006 cannot be taken as an indicator of the prevalence rates for the period prior to onset of MS. Therefore, if the effects of internal migration on the aetiology of MS are to be examined more fully then further analyses must be carried out to establish prevalence rates for the period prior to onset of MS.

There does appear to be month of birth effects in the prevalence of female RR/SPMS cases which are strongest for more southerly latitudes. Specifically, there appears to be a beneficial effect of mid summer a month or two before birth, and a non-beneficial effect of mid winter half way through the pregnancy. There are also competing effects for male RR/SPMS cases: a beneficial effect at the end of the second and start of the third trimester which is apparent in the north; and a non-beneficial effect of mid winter around birth in the south. There may be other effects present for other categories of MS cases, however, low sample numbers preclude any firm conclusions.

[This page intentionally left blank.]

Chapter 6

Variations in the Prevalence of MS with UV Exposure

6.1 Introduction

As discussed in Chapter 3 and demonstrated in Chapter 5, there is a well established relationship between prevalence of MS and latitude. It is also well accepted that latitude is a surrogate for some other factor (or factors) with exposure to UV radiation usually considered to be the most likely factor involved, via vitamin D synthesis, as described in Chapter 2. This Chapter examines the New Zealand UV data which is available, and the sunlight exposure data which was collected as part of the NZMSPS project. Note that the study covered by this dissertation is part of this larger project. As such, a thorough, in depth investigation into the relationships between the MS sun exposure data and UV exposure data was not possible. However, this Chapter will provide a discussion of analysis techniques and of the available data as an aid to further work.

Section 6.2 first looks at spatial and temporal variations of solar and UV radiation and why exposure varies with latitude and season. Different components of solar radiation, including UV, are then described along with action spectra. The section concludes with details of the UV data available from New Zealand's National Institute of Water and Atmospheric Research (NIWA). Section 6.3 examines the relationship between average UV exposure and latitude and discusses whether latitude could be considered as a surrogate for UV exposure. Proposals for further analysis and limitations of the data are also discussed.

6.2 Variations in Solar and UV Radiation

Spatial and temporal variations exist for solar radiation—the further away from the equator one travels the colder it gets (with less 'warmth' in the sun), an effect which is also seen with the coming of the winter season along with a shortening of the length of day. Further details of these variations will be given in section 6.2.1 as an aid to understanding the UV data which is then presented in section 6.2.2.

6.2.1 Insolation and UV Irradiance

The principles involved in understanding the spatial and temporal variations of solar radiation exposure (insolation) are similar regardless of which type (wavelength) of radiation is being considered. Therefore, the general principles for all solar radiation will first be outlined, followed by considerations which need to be made for different wavelengths. Note that a full description of the atmospheric physics and chemistry is well beyond the scope of this dissertation, though some further detail on theoretical exposure levels is given in Appendix F.

A recent review article which summarises solar radiation and many of the factors involved can be found in Juzeniene et al (2011). The main latitude effect is due to absorption and scattering of radiation in the Earth's atmosphere. In essence, the lower the sun is in the sky the more atmosphere the solar radiation has to pass through before reaching the ground, and so more radiation is absorbed; very little UVB gets through at lower solar elevations. In addition, higher levels of UV are experienced at higher altitudes since the solar radiation passes through less atmosphere.

The solar radiation incident on a flat surface is determined by the angle of incidence of the sun to that surface. As such, it depends on the angle of that surface to the Earth's surface and also the angle of the sun relative to the Earth's surface. The angle of the sun, in turn, depends on the time of year (and day) and the latitude of the surface concerned. If the Earth had no atmosphere then the exposure on a vertical surface in winter could match that of a horizontal surface in summer; the presence of the Earth's atmosphere means that, in practice, this does not occur. However, in winter, solar radiation incident on a vertical surface could be several times that which is incident on a horizontal surface. Since measured solar (and UV) exposure data is given in terms of exposure on a horizontal surface (irradiance), if the sun is not directly overhead then any surface which is not horizontal could experience a greater exposure than that given by the measured data. This observation is often overlooked in health research (Juzeniene et al. 2011; Moan et al. 2008). As Juzeniene et al point out:

"Calculations of UV exposure to human skin are usually represented by irradiances. This implies that the skin is represented by a plane, horizontal surface. This is obviously a bad approximation, and we have performed more realistic calculations by modelling the skin as a vertical cylinder" (Juzeniene et al. 2011, p.11).

Changes in exposure to UV due to the orientation of the surface exposed have also been investigated by others, including Webb et al (2011a) (discussed in section 2.2.4.2) and Kimlin et al (2003) who considered erythema (sunburn) and vitamin D production in the human face. Note that although considering an exposed surface as vertical rather than as horizontal would increase exposure levels in winter months, it would also reduce exposure levels in summer months, all other factors being equal. Similarly, exposure levels at noon would decrease and exposure levels at times other than noon would increase. In addition to the annual cycle of exposure there is also a daily cycle from sunrise to sunset. The first observation for the daily cycle is that there is considerable variation in the exposure levels throughout the day—sunrise and sunset do not simply 'turn on' and 'cut off' a fixed level of exposure. Rather, these limits are superimposed upon similar cycles to those seen for the annual variations, including variations due to the angle of the surface. A further complication to consider is that of human behaviour: people move around. This varies the body area presented to the sun such that a fixed angle surface model (whether horizontal or vertical) would not be appropriate—exposure to UV will continually change within the limits of the solar cycles described above.

Although the factors which affect incident solar radiation are known, combining them to predict incidence at ground level is not trivial, especially if random weather factors such as cloud cover must also be taken into account. In addition, the preceding descriptions only consider general solar radiation; factors such as absorption through the Earth's atmosphere also depend upon the chemical mix of gases present and the wavelengths of the radiation being considered. Figure 2.8 illustrated different wavelengths of solar radiation with respect to absorption through the skin. In a similar way, different gases in the atmosphere absorb different wavelengths of light by different amounts. In contrast, water droplets in clouds do not absorb UV radiation, however they do have a significant effect on ground level radiation levels by scattering radiation rather than by absorbing it. Figure 6.1 gives typical UV radiation levels which reach ground level for different solar elevations, and cloudless and cloudy conditions. These spectra were generated locally from Engelsen's FastRT simulation tool (Engelsen 2010a; Engelsen and Kylling 2005) which can be downloaded and run on Linux or a Windows PC. It can be seen that radiation levels decrease as the sun moves from overhead (solar elevation 90°) towards the horizon. In addition, the presence of clouds

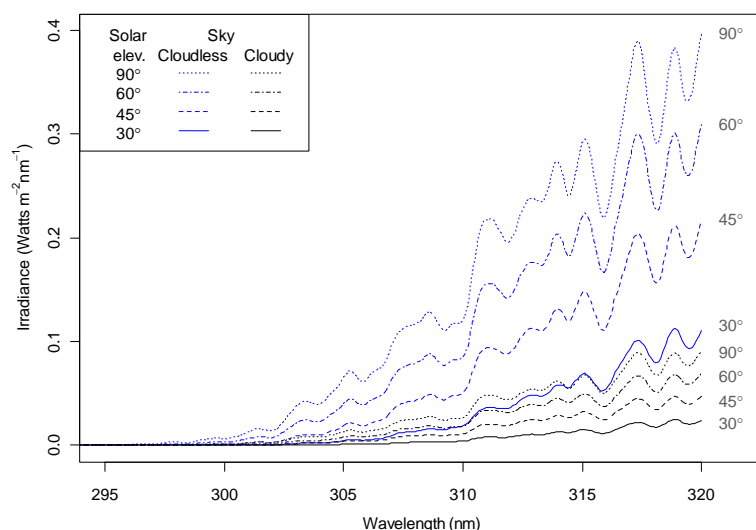


Figure 6.1 Simulated UVB radiation—solar elevation and cloud

UV radiation simulated for different solar elevations and for cloudless and cloudy sky. Data was generated using the Linux version of FastRT downloaded from the web site <http://nadir.nilu.no/~olaeng/fastrt/fastrt.html> (Engelsen 2010a). Solar elevations are as given in the key, and to the right of the spectra. Settings for cloud and ozone were the defaults used by the online version, as follows. All skies: ozone 350DU, albedo 0.03, surface altitude 0.15km, triangular spectral response function 0.6nm FWHM. Cloudless skies: visibility 25km. Cloudy skies: cloud liquid water column 400g/m².

reduces the levels considerably. For the default cloud settings used here, the highest UV levels are similar to those obtained for a low sun in cloudless conditions. It is also noticeable that there is very little UV at wavelengths less than 300nm.

In order to estimate solar radiation exposure, and especially UV exposure, a purely modelling approach is unreasonable for the current study, so another approach is required. Measured data, and modelled data based on the measured data, is available through New Zealand's National Institute of Water and Atmospheric Research (NIWA), which is described in section 6.2.3. First it is convenient to provide a description of action spectra, which indicate how important different wavelengths are when considering different chemical and biological reactions.

6.2.2 Action Spectra

Different wavelengths of radiation can have different effects when involved in chemical or biological reactions and the degree to which different wavelengths cause a reaction is termed the action spectrum. When considering a dose of radiation in a particular context the relevant action spectrum is used to weight the incident radiation. Unfortunately, although the functional forms of many action spectra are

known in principal, often the detail is allusive, being hard to obtain from experiment, or prone to differ between *in vitro* and *in vivo* derivations. In addition, as detailed in section 2.2.4.1 for vitamin D, the synthesis of a specific photoproduct is only part of a complex equilibrium with many factors playing a role. As such, a single wavelength-dependent action spectrum for the given photoproduct must be a compromise. Further, there has recently been a recommendation that the vitamin D spectrum should be revisited and reconstructed using more up to date techniques in order to address some of the factors which are now seen to be involved with vitamin D synthesis (Norval et al. 2010).

Action spectra for erythema (Webb et al. 2011b) and vitamin D synthesis (CIE 2006; Engelsen 2010a; Webb and Engelsen 2006) are illustrated in Figure 6.2 (a) and (b) along with a (smoothed) typical solar UV radiation spectrum (Kerr 2005). Dosages, which indicate how effective the incident radiation is, are determined as the convolution products of the solar spectrum with the action spectra, and are shown in (c). Note that there are two vitamin D spectra given, with the older 2005 vitamin D spectrum shifted by a few nm from the newer 2006 spectrum. However, both spectra are based on the action spectrum from MacLaughlin et al (1982): the '2005' spectrum is the "Vitamin D action spectrum (old version)" obtained from the FastRT online simulation tool (Engelsen 2010a); the '2006' spectrum is the "Vitamin D action spectrum" from the same FastRT tool, and is taken from the CIE Technical Report (CIE 2006) which defines the standard for the action spectrum. (Here, 'vitamin D spectrum' without regard to year or 'old' will imply the newer 2006 spectrum.) The note following Table 1 of the CIE Report, which tabulates the vitamin D action spectrum, is worth repeating here:

"In digitizing and tabulating this action spectrum we identified a potential source of confusion, illustrated by the action spectra derived by several independent groups falling into two sets, separated by several nm. This appeared to arise from the use of reproduction copies of the original manuscript as a source for the action spectrum. In photocopying or scanning, figures are not always reproduced exactly. Further curve tracing or manual extraction of data points then propagates the reproduction error. Figures in this report have been derived from an original manuscript." (CIE 2006, p.7)

Although the erythemal and vitamin D action spectra appear similar, apart from a shift of both vitamin D spectra to slightly longer wavelengths, once an incident UV spectrum

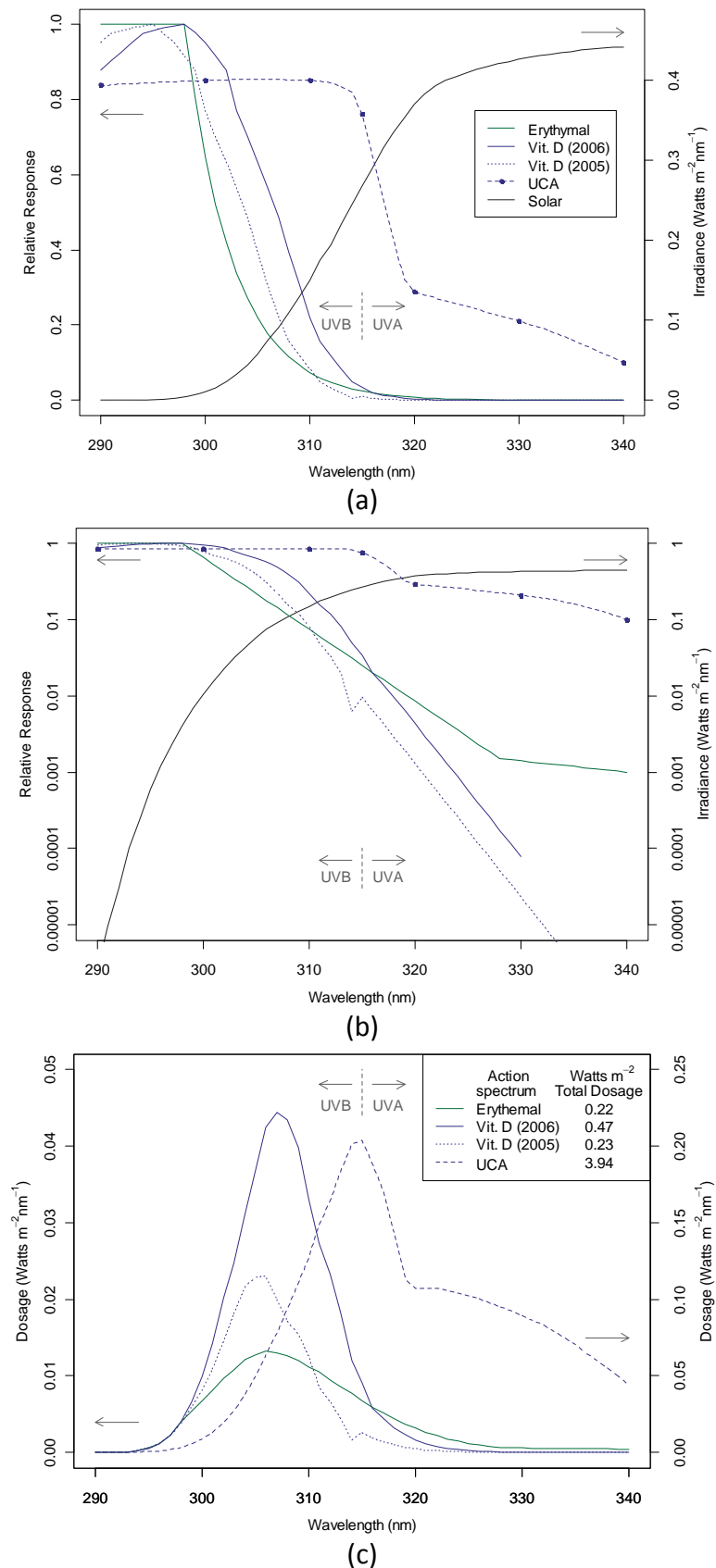


Figure 6.2 Action Spectra and Solar Radiation

Spectra for erythema, vitamin D and urocanic Acid (UCA) on **(a)** linear scale and **(b)** log scale (key: same as (a)). **(c)** The product of the action spectra with a smoothed version of the solar UV spectrum for a summer day in Toronto (43.7°N) (Kerr 2005). Action spectra: Erythymal: Webb et al (2011b); Vitamin D: 1987: Engelsen (2010a); 2006: CIE (2006); UCA points: McLoone et al (2005); UCA curve: interpolated for this figure.

is taken into account, there are noticeable differences in the dosages: the maximum vitamin D dosage is over three times that of the erythemal maximum dosage, and the total vitamin D dosage (obtained by integrating over the spectrum) is double the total erythemal dosage. The 2005 vitamin D spectrum is closer to the erythemal spectrum and gives only a slightly higher total dosage. Any studies which are based on this older action spectrum should probably be revisited, and results would not be comparable with results based on the 2006 CIE action spectrum, as mentioned by McKenzie et al (2009). Also note that the CIE spectrum is only defined to 330nm. In addition, vitamin D production is negated by reverse reactions above 315nm, so although the action spectrum indicates an effective dose above this wavelength, in practice there is a cutoff at 315nm (CIE 2006; MacLaughlin et al. 1982).

In section 2.2.4.2, vitamin D synthesis from UV exposure, in units of SED, was equated to an equivalent oral intake of vitamin D, and 1.1 SED was found to be equivalent to 2000 ± 600 IU of oral vitamin D when 35% body area was exposed to UVR. It was noted that $1 \text{ SED} = 100 \text{ Jm}^{-2}$ of erythemal UVR (UV_{Ery}). From Figure 6.2(c) it can be seen that the vitamin D effective UV (UV_{VitD}) dosage is 2.14 times that of the UV_{Ery} dosage, for the example solar spectrum in Figure 6.2. This suggests that around 235 Jm^{-2} of UV_{VitD} would be required to synthesise the equivalent of 2000 ± 600 IU of oral vitamin D, under similar conditions. If this was considered as an average daily dose, it would amount to around 7.15 kJm^{-2} of UV_{VitD} per month.

In Chapter 2, urocanic acid (UCA) was mentioned as an influence in the immune system which photosynthesises UV radiation. An action spectrum for UCA is also given in Figure 6.2 based on Figure 2 of McLoone et al (2005). Most notable for UCA is that its action spectrum is shifted to longer wavelengths than the erythemal and vitamin D spectra and it appears to have much more of a tail through the UVA wavelengths and towards the visible. As a consequence, the peak dosage for UCA is more than an order of magnitude greater than that for erythema and 4–5 times that for vitamin D. In addition, subject to uncertainties in the UCA action spectrum, and variations in the solar spectrum, the UCA dosage from UVA (2.5 Wm^{-2}) is greater than that from UVB (1.4 Wm^{-2}).

As discussed in Chapter 2, there is still uncertainty as to which biochemical reactions are important when considering the role of UV in the aetiology and pathogenesis of

MS. However, considering the relative effects of UVA and UVB as described above, and the greater penetration depth into the skin of UVA (Figure 2.8), studies looking at the effects of UV should not be restricted to UVB but should also consider the effects of UVA.

Now that aspects of UV radiation have been described, the next section looks at empirical data which is available for New Zealand.

6.2.3 NIWA UV Data

New Zealand's National Institute of Water and Atmospheric Research (NIWA) provide "UV Atlas" PC software (NIWA n.d.) which can be freely downloaded from the UV Atlas web site:

NIWA: <http://www.niwa.co.nz>

UV Atlas:

<http://www.niwa.co.nz/our-services/online-services/uv-and-ozone/uv-atlas>

Once downloaded and installed, this software will download and then process data to produce maps and time series data for UV irradiance around New Zealand. Most of the data is based on the locations of weather stations which provide measured broadband radiation data; these are distributed across New Zealand, as shown in Figure 6.3. Data is generally available from 1960 through to the current time, though availability varies considerably from station to station with many stations only having data from much later than 1960, for example, from 1978 or 2002. A brief guide outlining the data which is available from the UV Atlas follows. However, for more technical details, the definitive descriptions of the data should be consulted which are available from the UV Atlas web site and also through the help system which comes with the software.

A summary of the data inputs to the NIWA UV Atlas, including the temporal and spatial resolutions of the data, are given in Table 6.1. The 'variable' spatial resolution of the NIWA data refers to the distributed nature of the stations providing the data. Data outputs are either in the form of interpolated surface plots (maps) which cover the extent of New Zealand, or time series which are for individual stations. Many of the output measures are available in either map or time series form, whereas others are only available in one form or the other.

In addition, a number of UV Irradiance weighting functions (action spectra) can also be used to modify the derived measures. These weighting functions are applied to the measures to filter out certain ranges of UV and give an indication of effective UV, or dose, as described in the discussion of action spectra, above. The provision of pre-weighted measures removes the need for the full spectra to be provided. (However, this also means that the user is limited to output based on the weighting functions provided.) When choosing which measure to plot within UV Atlas, a choice is given of the weighting functions given in Table 6.2.

The default option is the Erythemally Weighted UV which applies the erythemal action spectrum to give a measure of 'sunburning' UV radiation. (In this chapter, 'erythemally weighted' will be assumed unless stated otherwise.) The UVI is simply a scaled version of this and gives a standardised index which is reported to the public. Note the two different wavelength ranges for UVB: 280–315nm and 280–320nm. The World Health Organisation (WHO) have specified the 315nm limit for UVB, however some fields of research still use the 320nm limit (Juzeniene et al. 2011), so both options are provided. Of most interest here is the weighting function for vitamin D synthesis. (Note that it would appear that none of these weighting function cover the UVA range. However, the Plant, DNA Damage and Cancer weighting functions would need to be investigated further before it can be ascertained whether or not this is truly the case.)

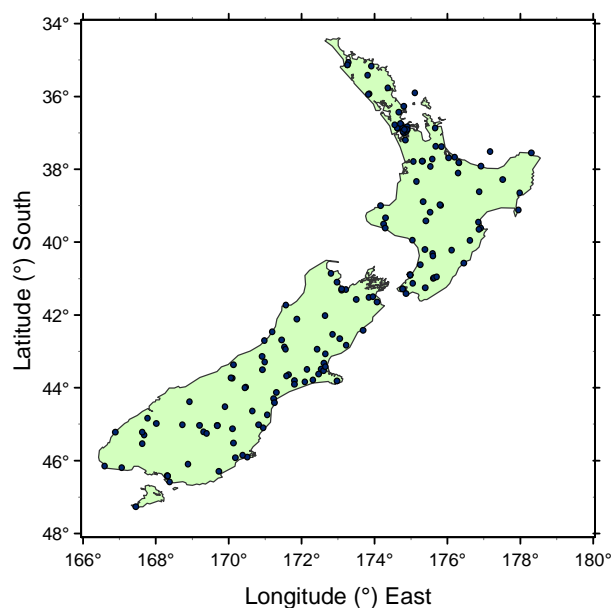


Figure 6.3 NIWA Weather Stations

The locations of NIWA weather stations around New Zealand for which measured broadband radiation data is available

Measure	Frequency	Spatial Resolution	Source
Measured Broadband Radiation	Hourly	Variable	NIWA ¹
Seal Level Pressure	Hourly	Variable	NIWA ¹
Temperature	Hourly	Variable	NIWA ¹
Humidity data	Hourly	Variable	NIWA ¹
Cloud Cover	Hourly	Sparse	NIWA ¹
Station Locations		Variable	NIWA ¹
Sea Level Pressure	12 hourly	2.5°x2.5°	NCEP/NCAR ²
Ozone data	Daily	1° lat x 1.25° lon	Combined ³
NZ Topography – altitude (m)		~1km	

¹ NIWA climate database stations; ² NCEP/NCAR reanalysis project;

³ TOMS (4 instruments), GOME, and OMI, homogenized and gridded

Table 6.1 NIWA UV Atlas Program—Data Inputs

Data inputs for the NIWA UV Atlas program are given with indications of the temporal and spatial resolutions of the available data.

UV Index (UVI)

Erythemally Weighted UV (McKinlay & Diffey) [Default]

Plant Damage Weighted UV (Caldwell)

DNA Damage Weighted UV (Setlow)

Plant Damage Weighted UV (Flint and Caldwell)

USC Skin Cancer (de Gruijl et al)

UVB 280–315nm

UVB 280–320nm

Vitamin D synthesis 320nm cutoff (MacLaughlin et al)

Table 6.2 NIWA UV Atlas Program—Weighting Functions

The weighting functions for the NIWA UV Atlas program. Names given refer to references given in the UV Atlas documentation.

Once plotted within the UV Atlas program, the derived data can be extracted from the surface plots or the time series and saved on the PC for use elsewhere. These outputs are given in Table 6.3 and will now be discussed in more detail. Surface plots and time series are discussed separately, partly for convenience here, and also because they are treated differently within the UV Atlas software and have different limitations.

	Hourly	Daily	Monthly	Yearly	Measured Broadband	Sea Level Pressure	Temperature	Humidity	Cloud Cover	Station Locations	Ozone	Topography
(a) Surface Plots												
Clear Sky Irradiances	✓	✓	✓	✓		✓					✓	✓
Estimated True Irradiances	✓				✓	✓	✓	✓	✓		✓	✓
Sea Level Pressure	✓	✓	✓	✓		✓						
Surface Pressure	✓	✓	✓	✓		✓						✓
Total Column Ozone		✓	✓	✓							✓	
Topographical map												✓
Locations map										✓		
(b) Time Series												
Clear Sky UV Dose	✓	✓	✓	✓		✓					✓	✓
Estimated True UV Dose	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
Cloud Cover Mod. factors	✓	✓	✓	✓	✓				✓			
Modelled Broadband Rad.	✓	✓	✓	✓		✓	✓	✓				
Measured Broadband Rad.	✓	✓	✓	✓	✓							
Sea Level Pressure	✓	✓	✓	✓		✓						
Surface Pressure	✓	✓	✓	✓		✓						✓
Temperature	✓	✓	✓	✓			✓					
Humidity	✓	✓	✓	✓				✓				
Total Column Ozone		✓	✓	✓							✓	

Table 6.3 The NIWA UV Atlas Program—Data Outputs

The data outputs from the NIWA UV Atlas program indicating the temporal resolutions of the data and the input data required for (a) surface plots and (b) time series.

6.2.3.1 Surface Plots

Table 6.3(a) summarises the surface plots available from the NIWA UV Atlas program. These plots are obtained by interpolating the data provided by the NIWA weather stations to a grid and then contouring the data for display in the UV Atlas viewer. Although the resolution of the data extracted from the surface plots is lower than that displayed by the UV Atlas program (which is determined by the viewer window), the 0.25° resolution which is provided is accurate enough for this study, especially considering the various approximations being made at other stages in the processing. However, the 'Estimated True Irradiance' measures and other derived data which take

into account cloud cover are only available for Hourly surface plots. (Unfortunately, this includes vitamin D and other weighting functions.) Data could be extracted for a very large number of Hourly plots and then averaged to get Monthly data. However, the plotting and extraction of the data is a manual process and the number of plots which would be required would make this approach unrealistic. An alternative approach, which makes use of clear sky irradiance maps in conjunction with time series data will be described below.

Interpolated maps of erythemally weighted clear-sky irradiances, averaged for each month over the five year period 1980–1984, are shown in Figure 6.4 as 'thumbnails' along with the average for all months. It is noticeable that local spatial variations are less than seasonal variations. Although the surface maps cannot provide an average of the estimated true UV dose data at the temporal resolution required, the time series can. The time series data which is available from UV Atlas will now be discussed.

6.2.3.2 Time Series

Table 6.3(b) summarises the time series available from the NIWA UV Atlas program. In contrast to the surface plot data for which the Estimated True Irradiance measures were only available Hourly, the Estimated True UV Dose data is also available at Daily, Monthly and Yearly resolutions.

Note that the Estimated True Dose data which is available is not complete for each station. For example, many stations have little or no data for June (mid winter) when UV levels are at their lowest and the monitoring equipment is likely to be taken offline for maintenance. In addition, some stations have prolonged periods of data unavailability between periods of almost constant data availability. The 22 stations for which most data was available are shown in Figure 6.5 and listed in Table 6.4. (These stations will simply be referred to as 'the 22 stations' in the rest of this section.) Also given in the table is an indication of the data which is available. Despite being in this list of 22 stations, those at Auckland and Leigh show only 88% and 70% of days covered, respectively, due to prolonged periods of data unavailability. In the case of Auckland no data was provided from May 1993 to July 1994. In the case of Leigh, the station provided no data from April 2009 to March 2010 or from January 1987 to December 1992. If these three prolonged periods are allowed for then Auckland and Leigh offer 93% and 95% data availability, respectively. As well as totally missing data,

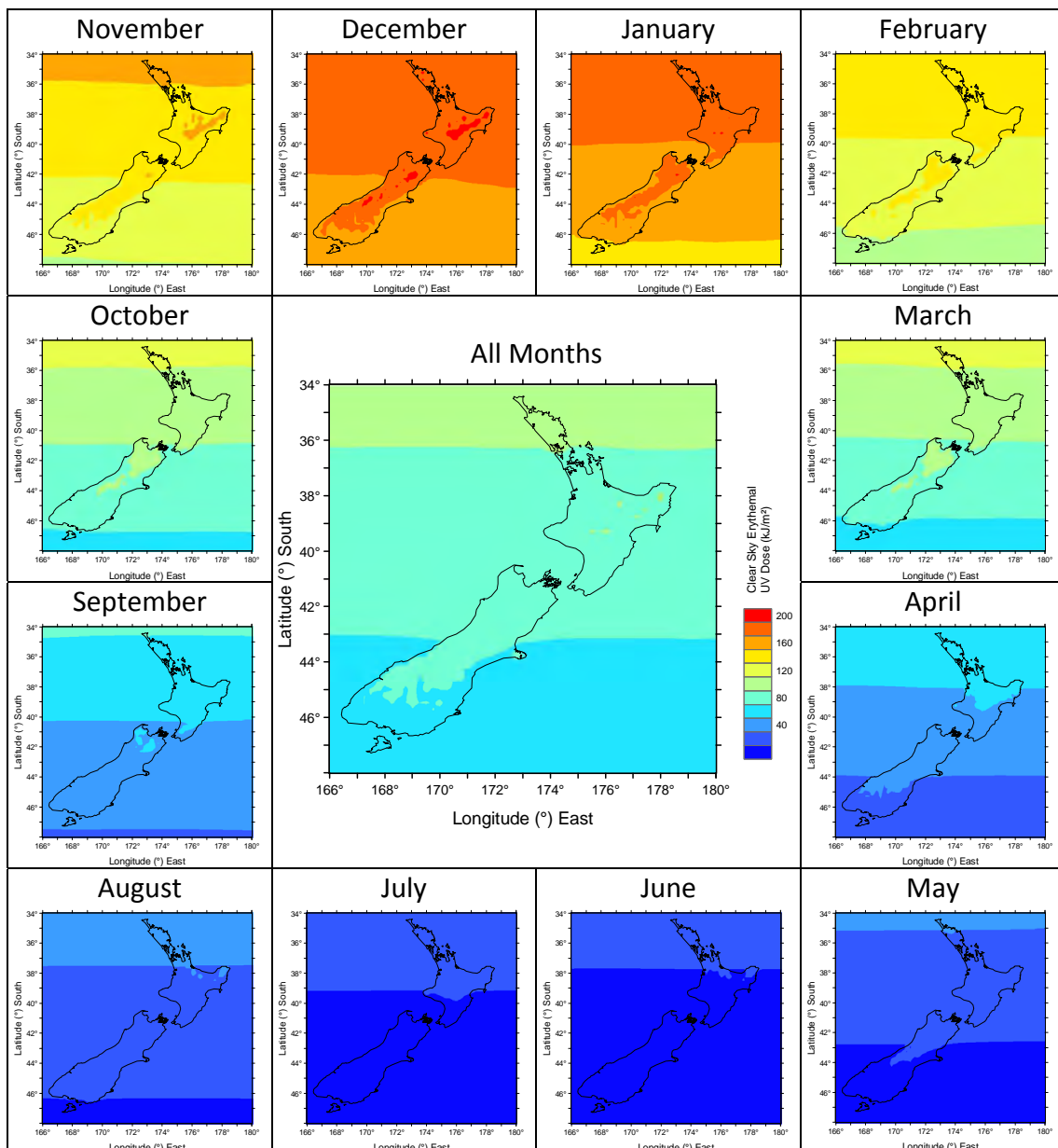


Figure 6.4 Spatial Variations in Clear-Sky UV by Month

Clear Sky UV Dose (Erythemally Weighted) for each month and for all months averaged over 1980–1984. The data range and colour scale are the same for all months.

data in some time periods may be low due to partially missing data (especially in those adjacent to others with missing data). For example, no data was provided from the Lauder station for the three days from the 3rd to the 5th January 1992. However, the previous hourly data was for 15:00 on the 2nd, and the following was for 11:00 on the 6th, which compare with 19:00 and 06:00 for the same days on other years.

The number of hours (or days) of missing data to allow before disregarding a day (or month) when aggregating data can be adjusted within the UV Atlas program. However, there is a compromise between data quality and quantity—restricting this to increase

the quality of the data also restricts the amount of data available. Note that night periods are treated as missing data, so the number of missing hours to allow must be greater than the length of a New Zealand night in winter if all year round data, or data for multiple years, is required. For the case of Lauder, the mid winter nights (no data) are 17 hours long so the number of missing hours has to be set to at least 17 hours to get any mid winter data from the UV Atlas program. However, applying the same limit to January, as with the missing data period mentioned above, then mid summer days of 14 hours could lack half of their data and still not register as missing. There is no easy way to gauge the extent of missing data, or the effect it might have on aggregate data. The UV Atlas program will report the hours of available data, however, this can only be done for one data output, one station and one year at a time.

To give an overall, albeit crude, measure of data availability, the data coverage figures presented in Table 6.4 were determined from Hourly Vitamin D weighted time series data downloaded from the UV Atlas program and then analysed using the statistics program R. Although the method used simply determined the coverage extent of the data for the 22 stations, without regard for any periods of data unavailability in between, it does give a reasonable indication of the proportion of days for which data is available. The prolonged periods given as examples above were picked out of the data by eye once the lower percentage coverage figures were noted. If a more detailed

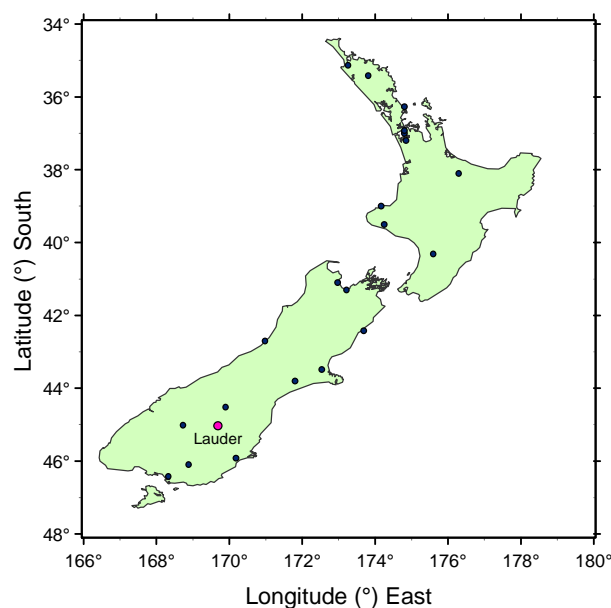


Figure 6.5 NIWA Weather Stations (22)—Locations

The locations of the 22 NIWA weather stations around New Zealand for which the most UV data is available. The location of Lauder, the base of NIWA, is indicated.

Station	Period Covered		Days Covered		
	From	To	Total	Data	%
Auckland Aero	01/1985	03/2011	9586	8482	88%
Auckland RC, Onehunga	10/1994	03/2011	6026	5842	97%
Christchurch Aero	01/1985	03/2011	9586	8866	92%
Dunedin Aero AWS	12/1985	03/2005	7061	6905	98%
Gore EDR	07/1985	03/2010	9040	8262	91%
Hokitika AWS	10/1991	03/2011	7122	6969	98%
Invercargill Aero	01/1985	05/2000	5630	5468	97%
Kaikohe EDR	12/1985	02/2011	9221	8339	90%
Kaikoura AWS	10/1991	03/2011	7122	6977	98%
Kaitaia EWS	01/1999	02/2011	4442	4250	96%
Lauder EWS	08/1985	03/2011	9374	8960	96%
Leigh 2	01/1985	03/2011	9586	6667	70%
Motueka, Riwaka EWS	06/1996	03/2011	5417	5290	98%
Nelson AWS	12/1991	03/2011	7061	6886	98%
New Plymouth AWS	11/1991	03/2011	7091	6950	98%
Normanby EDR	04/1986	07/2004	6697	6118	91%
Palmerston North AWS	12/1991	03/2011	7061	6933	98%
Pukekohe EWS	03/1986	03/2011	9162	8730	95%
Queenstown AWS	10/1991	03/2011	7122	6941	97%
Rotorua Aero AWS	11/1991	03/2011	7091	6943	98%
Tara Hills EDR	04/1985	03/2011	9496	9014	95%
Winchmore EWS	03/1996	03/2011	5509	5424	98%

Table 6.4 NIWA Weather Stations (22)—Data Availability

Data availability for the Hourly Total Estimated True UV Dose (Vitamin D weighted) for the 22 NIWA weather stations around New Zealand for which the most UV data is available. Availability is based on the presence of data for 12:00, midday.

and systematic exploration of data availability was required then further analysis of Hourly data using R (for example) would be recommended, rather than an interactive analysis through UV Atlas. The effect of missing hours or days of data is likely to be minimal for the analysis, described below; monthly totals which are affected will be outliers below the main body of data. Given the bulk of data, however, any shift in the trends observed will be well within any uncertainties due to other factors.

Time series data is available as monthly totals for the weather stations around New Zealand as indicated in Figure 6.3, though for this analysis only data for the 22 stations described above was used. The clear-sky irradiances were downloaded as monthly

totals for each of the 22 stations, along with the Estimated True UV dose data weighted by Vitamin D and the two UVB weighting functions. (Here UVB1 will be used to designate the weighting for the 280–315nm range and UVB2 will designate the weighting for the 280–320nm range.) Figure 6.6 shows the weighted UV data plotted against the clear-sky data for the Lauder station and for all 22 stations considered in this analysis. The data for Lauder highlights the grouping of the Monthly totals, with the data from the two peak summer months, December and January, being distinctly separate from the next two months, February and October. A similar pattern is seen right through the year with the four main winter months, May through to August, being grouped with the lowest doses. Also of interest is that, as well as being separated in terms of the measured clear-sky dose, the groups are also almost totally separated in terms of the modelled, true weighted dose—the variability in the Monthly totals from year to year is less than the variability from month to month, apart from mid winter and mid summer. Once all of the stations are grouped together, however, the monthly divisions merge due to the variations in irradiance from the spatial distribution of the stations. It is also interesting to note that the data is not heteroscedastic unless plotted on a log–log scale.

The relationships between the monthly totals for the weighted estimated true UV doses and the clear-sky doses are generally linear. The log–log plots suggest that there is some non-linearity at low doses, though whether this is a true effect or whether it is an artifact of missing data is uncertain. Based on mappings such as those illustrated, it would be possible to estimate an average weighted true UV dose from the clear-sky dose. However, a similar investigation of this mapping for various locations in the United States by Pope et al (2008) suggests that the mapping may depend significantly upon the time of day. A breakdown of the mapping to this degree of detail would not be feasible here. In addition, the MS case exposure duration data which is available for this project does not include any indication of time of day, so such detail would be superfluous in this context. The monthly totals can, however, be useful when considering an average UV dose in the context of vitamin D synthesis, as will now be looked at, briefly.

It was estimated in section 6.2.2 that for the example solar spectrum shown in Figure 6.2 for Toronto, (43.7°N), in mid summer, a monthly UV_{vitD} dose of around 7.15 kJm^{-2}

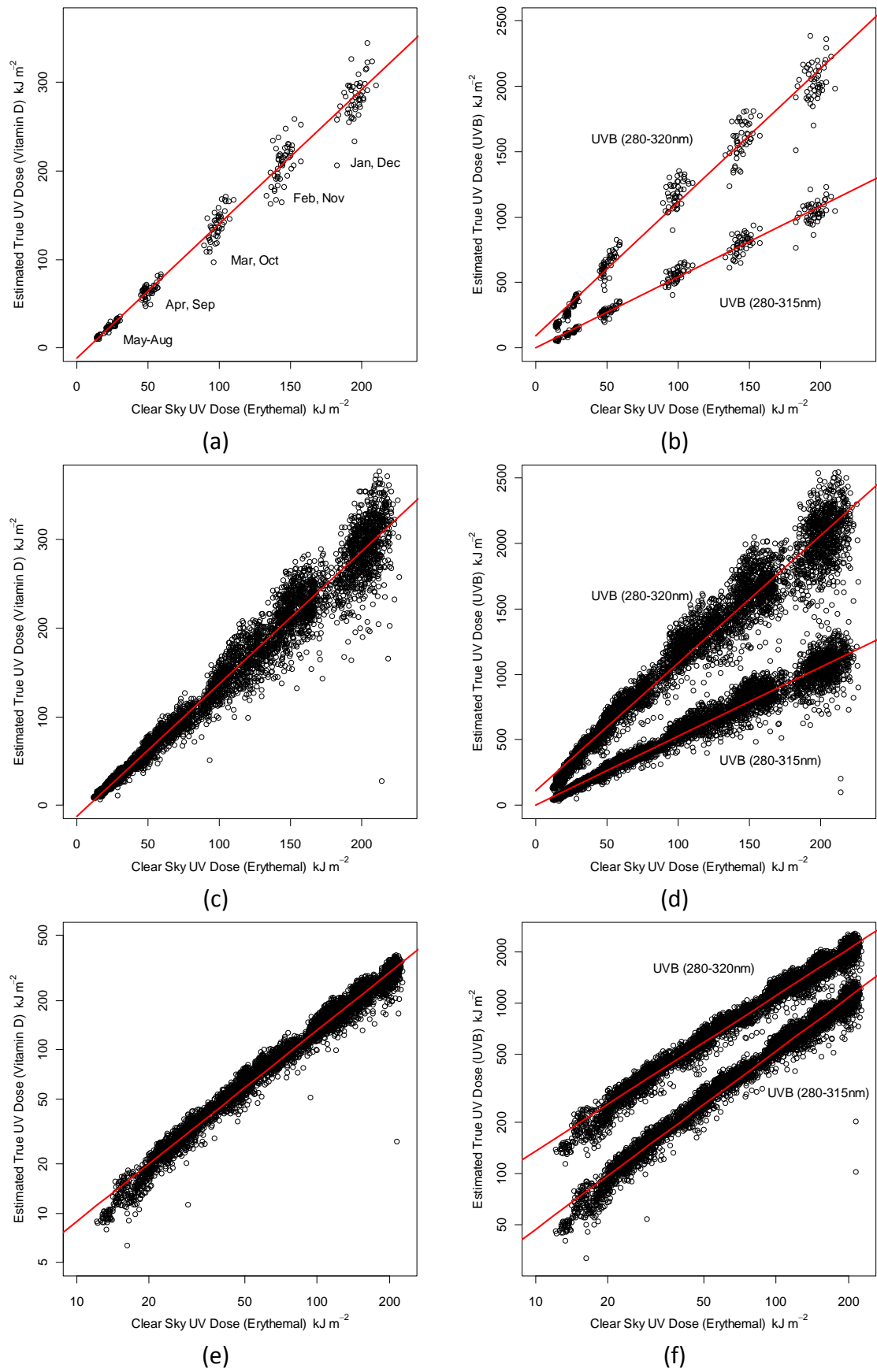


Figure 6.6 Time Series—Monthly Totals

Monthly totals (1985 to March 2011) for weighted Estimated True UV Dose against clear-sky erythemally weighted dose. UV Weightings: **(a)**, **(c)** and **(e)** Vitamin D; **(b)**, **(d)** and **(f)** UVB1 and UVB2; Stations: **(a)** and **(b)** Lauder; **(c)**, **(d)**, **(e)** and **(f)** all 22 stations. Red lines: regression fits: **(a)–(d)** linear; **(e)** and **(f)** log.

would be required to synthesise the equivalent of 2000 IU of oral vitamin D per day. Within the uncertainties of this example, Lauder (45°S) is likely to experience a similar solar spectrum on a summer's day, so the UV_{vitD} required for an equivalent oral vitamin D dose is also likely to be similar. From Figure 6.6(a), this should easily be achievable in summer. Assuming a similar figure for winter, however, suggests that a considerable proportion of time would need to be spent outdoors in order to achieve this vitamin D dose equivalent from UV exposure alone: the mean July monthly total for Lauder is only 12.5 kJm^{-2} of UV_{vitD} , less than twice the estimated required monthly dose. Although an upright body position will receive a higher dose for the lower solar angles experienced in winter, the estimated monthly dosage was also based on a 35% body area exposed, which is unlikely on a regular basis, especially in mid winter. Consequently, much of the required daily intake of vitamin D must come from diet over the winter months rather than from UV exposure alone. Although the FastRT tool used to generate the spectra in Figure 6.1 could be used to generate a collection of (simulated) solar spectra for different locations around New Zealand, and for different times of the year, in order to determine more detailed limits on vitamin D synthesis, this was not carried out at this time.

This section has established the availability of empirical UV data from NIWA. The next section will look at how that UV exposure data varies with latitude within New Zealand, before going on to examine the relationship of this data with the prevalence of MS.

6.3 Average UV Exposure with Latitude

Data extracted from the surface plots discussed in section 6.2.3.1 includes total monthly clear sky UV exposure. Average total monthly exposure can be determined, either by month, or for all months (Figure 6.4) by averaging this data over a number of years. This data can be aggregated further using the same latitude slices which were used in the prevalence–latitude analyses in Chapter 5. Figure 6.7 shows the population weighted average total monthly exposure against population weighted latitude for the 2006 census data and the 1956 (Age 0–4) census data. Although this UV data is clear sky, Figure 6.6 suggests that including an average mapping from clear sky UV exposure to estimated true UV dose would only affect the gradients and intercepts of the two regression fits in Figure 6.7; the relationships would remain linear.

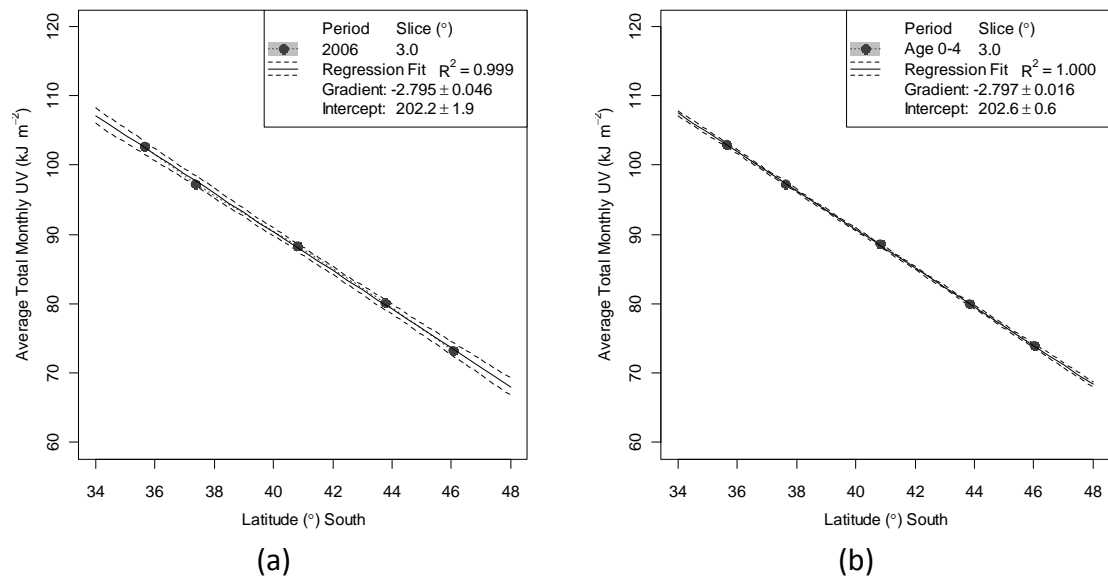


Figure 6.7 UV Exposure – Latitude

Average total monthly UV exposure with latitude. The UV exposure measure is clear sky erythemally weighted. Both measures are population weighted to **(a)** 2006 census data and **(b)** 1956 (Age 0–4) census data.

From Figure 6.4 and Figure 6.6, it is also likely that the relationships between average total monthly UV and latitude for individual months will be similar to those for the annual monthly averages and will only vary in the gradients and intercepts of the regression fits. However, there would be greater uncertainty in the data points and confidence bands would be wider due to the lower sample numbers after stratification by month. (These analyses, by month, were not carried out at this time.)

Figure 6.8 shows prevalence of MS against clear sky UV exposure. Note that the actual averaged UV data is used, not values determined from the regression fits from Figure 6.7. Given the linearity of the UV exposure with latitude (Figure 6.7), it is not surprising that the relationships between prevalence of MS and UV exposure shown in Figure 6.8 (a), (c) and (d) are almost mirror images of those shown in Figure 5.7(b), Figure 5.8(b) and Figure 5.9(b).

Although the UV exposure data illustrated here is an average of the total monthly clear sky data (erythemally weighted), this analysis demonstrates that latitude can be considered as a surrogate for average monthly UV exposure for the given latitude. It must be emphasised, however, that this average takes no account of any vitamin D weighting, nor any measure of the estimated exposure experienced by the MS cases. It would be interesting to see whether incorporating an estimate of actual exposure to

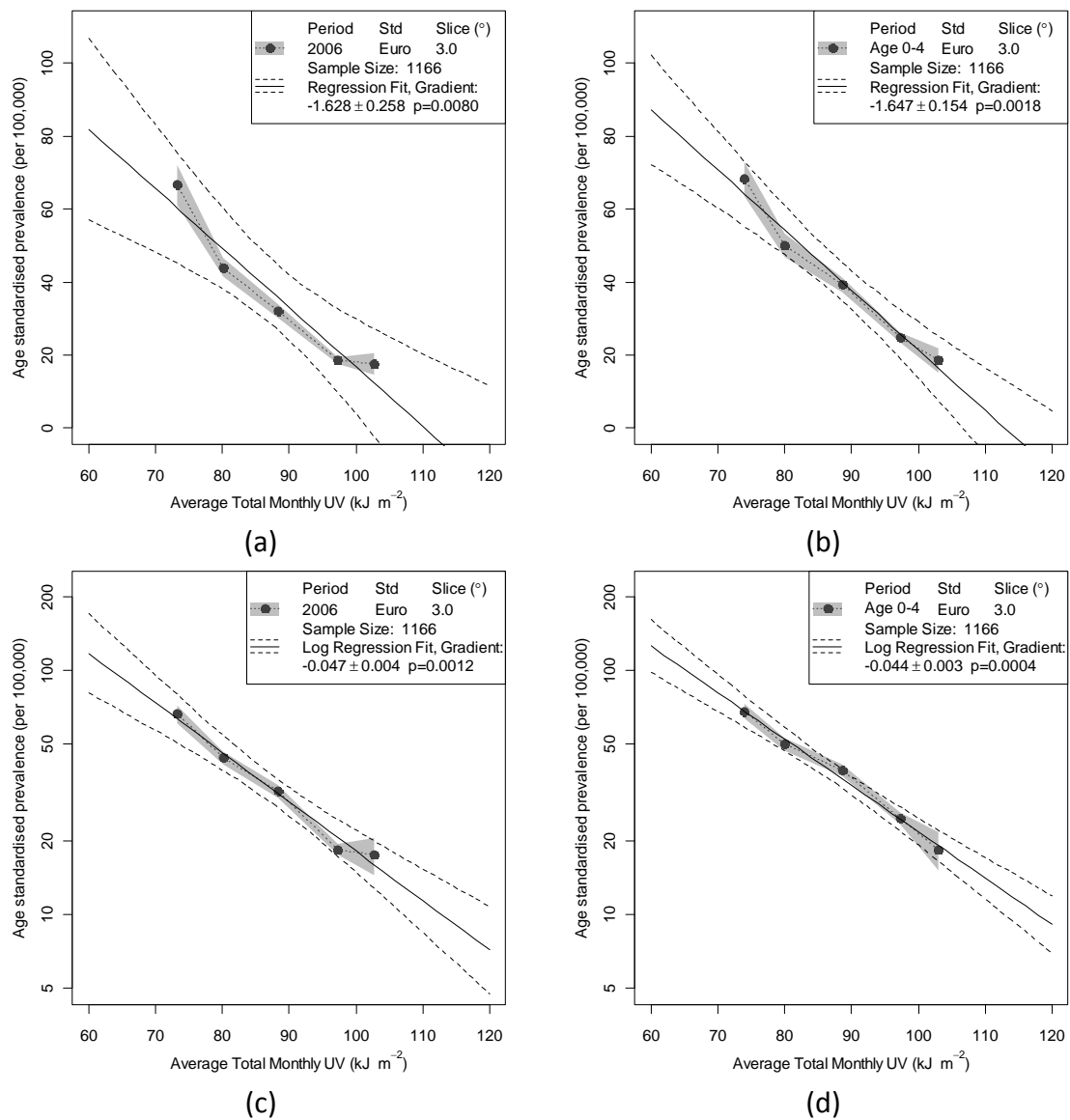


Figure 6.8 Prevalence of MS—UV Exposure

Age-standardised prevalence of MS in New Zealand with average total monthly clear sky UV exposure (erythemally weighted). **(a)** and **(c)** prevalence for census day 2006; **(b)** and **(d)** prevalence for residence location at Age 0–4. Regression fits: **(a)** and **(b)** linear; **(c)** and **(d)** log. Latitude and UV exposure are population weighted. (Slice size: 3°; Standard: European.)

sunlight from the NZMSPS survey data, strengthens the relationship observed in Figure 6.8, however, further analysis of the NZMSPS exposure data was not carried out at this time. These aspects will be discussed in the next section, 6.4.

6.4 MS Prevalence and UV—Discussion of Analysis

Although UV exposure is believed to be a surrogate for other factors such as vitamin D, as discussed in detail in section 2.2, the relationship between UV exposure and, for example, blood serum levels of vitamin D is not trivial (Engelsen 2010b). A study by McCarty (2008) looked at whether sunlight questionnaires could be used to assess vitamin D status. The main points of note here are that there was a low correlation between self-reported sunlight exposure and estimated UV dose, and also a low correlation between estimated UV dose and vitamin D status. In addition, Lucas et al (2011b) found that sun exposure and vitamin D blood serum levels were independent risk factors for first CNS demyelinating events (FDEs), a precursor to MS. Therefore, this discussion will concentrate on the analysis of UV exposure rather than any underlying biochemical factors.

In order to link prevalence of MS with exposure to UV two measures are required: the duration of exposure to UV and the dose experienced for that exposure. As detailed in section 3.3, the NZMSPS questionnaire asked respondents to indicate the average number of hours spent in the sun per week during summer and winter for each residence location. No details were asked for regarding time of day for the exposure, and there is also potential for misclassification due to recall bias—although these issues were also discussed in section 3.3, potential misclassification of this data, which might affect results obtained, will be examined towards the end of this section, once potential methods for analyzing the UV data are discussed.

The NZMSPS questionnaire data provides the duration of exposure to UV. However, as with all of the data within this study, there is no control data—the 'normal' durations for exposure to UV are unknown. (For looking at the prevalence of MS with latitude, it was sufficient to use the census population data to give the 'normal' population with latitude.) Therefore, before any final analyses and conclusions can be attempted, further investigation of various analytical and statistical methods for analysing the data must be made in order to establish which methods are actually suitable. Meanwhile, it is useful to discuss methods for determining estimates of the UV doses experienced by the MS cases. Note that here all data is weighted by one or other of the weighting functions so the term 'weighted' will be dropped as being understood in the current context. Also, the weighting functions apply the relevant action spectra inherently; it

is not possible to apply the required spectrum independently of the UV Atlas software since exposure data is not available as a function of wavelength.

An estimated true UV dose experienced for a given duration during a given month can be obtained by taking the clear-sky dose surface plots described in section 6.2.3.1 and mapping the data to the required true dose using the ratio of clear-sky dose to true dose, as described in section 6.2.3.2. The estimated true dose data obtained would vary by month of year and by location within New Zealand. As such, it could then be combined with the residence locations and the exposure durations from the NZMSPS data to give an estimated true dose for each MS case for the period of concern.

Assuming that there is a link with vitamin D, and noting from sections 2.2.4.2 and 6.2.3.2 that UV exposure in winter may not be adequate for maintaining an optimum vitamin D status, then estimates of seasonal doses should also be determined.

However, there are also a number of questions to ask in relation to the estimate obtained for the dose received. First, how good an estimate is this dose? Estimates of doses obtained are likely to be over-estimates for summer and under-estimates for winter, based purely upon angle of surface considerations. In winter, when people tend to wrap up against the weather, it is likely that as little as 10% of the body is exposed to sunlight, whereas in summer as much as 63% could be exposed regularly (McKenzie et al. 2009). (Though only half would face the sun at any given time, scatter would provide some exposure to the other half.) However, the amount of clothing worn does also depend on weather and other life style and social considerations. In general, considerations of clothing imply that estimates of UV doses will vary throughout the year with the greatest over-estimates in winter.

There is another consideration, as discussed in section 2.2.4.2—behaviour associated with protection from over exposure to UV. Although this is most likely to reduce summer UV exposure rather than winter exposure, it is still a factor which should be taken into account. However, if the period of concern in the lives of the MS cases is the Age 0–4 period, with a mean of 1954, then this may not be as important an issue as the other factors described above since there was not the same awareness for protecting against over exposure as there has been in more recent years. In addition, the use of sunscreen was not as widespread then as it is now, and the sunscreens which were used were not as effective as those recommended for use today. There is

also a further consideration with regards to urocanic acid (UCA) and sunscreens. After UCA was discovered to be a "natural sunscreen" in 1957, commercial cosmetic producers included it in sunscreen products until its immunosuppressive properties were discovered in 1983, and it began to be removed from products from 1991 (Gibbs and Norval 2011). Use of sunscreens up until the early 1990's may, therefore, be confounded by the inclusion of UCA.

All of the factors described above will confound any estimate of the true UV dose which the MS cases have been exposed to, whichever weighting function is used. Bias could be systematic, combining to produce a noticeable shift in the effects observed. However, it is also possible that these factors will broadly cancel out or not be relevant for the periods under consideration. Further investigations would need to be made into the scale of any effects before any conclusion on this could be made. Even if it is the case that the average UV exposure data determined in section 6.3 could be used to represent true exposure, after taking these confounding factors into consideration, it is then still necessary to weight the average UV exposure data by the estimated sunlight exposure data from the NZMSPS survey, as discussed above. Prevalence analyses based on this combined data could then be used to determine whether or not latitude can be considered to be a surrogate for estimated actual UV exposure when looking at the prevalence of MS within New Zealand.

Potential misclassification of the average UV exposure data and the sunlight exposure data must also be investigated further as part of any future analyses. However, a similar analysis of UV exposure data was made by Tatalovich et al (2006) in a case-control study looking at the risk of melanoma in Los Angeles County, USA. Although the results and conclusions of that study are not directly transferrable to the study of risk of MS, many aspects of the analyses are relevant since the authors examined the importance of: average annual UV exposure; estimated lifetime UV exposure; UV exposure at a young age; self-reported time spent outdoors; and the effect of overseas residency. Any misclassification of the UV exposure estimates, and the self reported time outdoors, for cases and controls was assumed to be non-differential, leading to a reduction in effects towards the null. However, it is also possible that a partially ecologic study such as this may see bias in either direction (Björk and Strömberg 2002; Jurek et al. 2005). (An ecologic study makes use of group or aggregated data such as

that obtained here through the GIS aggregation methods rather than individual case data.) Although similar conclusions on misclassification can be made for any analyses which might be made on the UV and sunlight exposure data discussed here, whether or not there is any control data will affect the results obtained. For a cohort study with no control data, any misclassification will be differential, leading to random effects; if suitable control data is made available for these analyses then the misclassification will be non-differential with a tendency for bias towards the null, which would lessen any effects observed. Further discussions on the effects of bias in ecologic and partially ecologic studies can also be found in the cited references and other sources (Björk and Strömberg 2005; Greenland 2001; Kim et al. 2010a; Thomas et al. 1993; Webster 2007).

6.5 Summary of UV Exposure

Prevalence of MS is negatively correlated with population weighted average total monthly UV exposure. The NIWA UV Atlas provides a source of UV data which can be used to investigate UV exposure. Estimates of average total monthly clear sky UV exposure doses for anywhere in New Zealand are available and time series data enables these clear sky doses to be mapped to estimated true doses. Monthly totals from the time series data suggest that, although UV levels are more than sufficient in summer for an individual to synthesize enough vitamin D for their requirements, it is unlikely that this would be possible in winter.

In addition to residence location data, the data available from the NZMSPS project provides an estimate of the time spent outdoors, in both winter and summer, by the MS cases under study in the project at any point in their life. This case data could be combined with the estimated UV data to estimate the actual UV exposure experienced by the MS cases. These estimated actual exposures could then be used in similar analyses to determine whether or not this strengthens the prevalence–UV exposure relationship. However, further analyses would also be required to establish whether bias introduced by confounding factors or misclassification of the UV or sunlight exposure data would be significant.

Chapter 7

Summary and Conclusions

7.1 Introduction

This Chapter summarises this study and its main findings. It will discuss implications of the results along with strengths and limitations of the study. Areas for further research will then be recommended along with concluding remarks.

7.2 Summary and Main Findings

The main aim of this study was to investigate the prevalence of MS within New Zealand, with particular reference to changes in prevalence with latitude of residence of the MS cases by gender, different MS phenotypes and different life periods. This study also determined the effect of different analysis techniques on the prevalence results obtained, and examined UV radiation exposure as a surrogate for latitude.

The data for this study was collected in parallel with the 2006 New Zealand population census as part of the New Zealand National MS Prevalence Study (NZMSPS) project, for which initial results have already been published (Taylor et al. 2010b). Since migration has been established elsewhere as a major influence on the effect of latitude on the prevalence of MS (Alter et al. 1978; Visscher et al. 1977), the full NZMSPS cohort of 2917 cases was divided in three groups. The first group consisted of 912 cases of MS for which actual residence location for the census day 2006 was either not known or not disclosed to this study, or where incomplete or no MS questionnaire data was returned to the NZMSPS project; this 'Missing' group was discarded for this study due to lack of data. The second group included all cases for which all residence locations were known and were within New Zealand from birth to onset of MS; this was the 'Resident' cohort of 1166 cases. The third group comprised the remainder of the cases, cases where a location was available for census day 2006 but for which at least one residence location between birth and onset of MS was not in New Zealand; this was the 'Migrated' cohort of 839 cases. The 'Combined' Resident plus Migrated cohort of 2005 cases was also considered for some analyses. Although the Migrated cohort had a slightly higher proportion of men than the Resident cohort (24.8% and 23.3%, respectively), the age profiles of the two cohorts were very similar. The prevalence of MS was determined for the Resident, Migrated, and Combined cohorts, stratified by

various factors. Māori ethnicity was not considered as a factor within this study due to insufficient cases of MS within the Resident cohort with self declared Māori ethnicity. This study established the following key findings.

7.2.1 General Prevalence

The prevalence of MS within New Zealand depends upon gender, MS phenotype, and the cohort of cases analysed. Although these factors are well established for the prevalence of MS, this is the first study to provide a detailed analysis of the prevalence–latitude gradient for these factors whilst controlling for the effects of external migration, or to look at the relationship between prevalence of MS and residence locations for ages 0–4 within New Zealand.

For the Resident cohort:

- the female to male (F:M) sex ratio is 4.03:1 for RR/SPMS and 1.55:1 for PPMS;
- 85.6% of female cases have RR/SPMS compared to 69.9% of male cases.

For the Migrated cohort:

- F:M is 3.54:1 for RR/SPMS and 1.58:1 for PPMS;
- 85.9% of female cases have RR/SPMS compared to 73.6% of male cases.

The prevalence of MS increases with increasing latitude south within New Zealand, though values obtained for the prevalence of MS vary depending on data aggregation methods used for the analyses. Here, latitude is taken as the population weighted centroid (PWC) of 3° slices of latitude, and prevalence is age-standardised to the European standard.

The prevalence of MS on census day 7 March 2006 follows a linear trend between the far north of New Zealand (PWC: 35.67°S) and the far south of New Zealand (PWC: 46.06°S), which is discussed further in section 7.2.2. The extremes for prevalence of MS on census day 7 March 2006 for the combined Resident and Migrated cohort were:

- for the far north of New Zealand 31.4 per 100,000 (95% CI: ±4.0);
- for the far south of New Zealand 100.1 per 100,000 (95% CI: ±6.7).

This is a factor of 3.2 increase from the far north to the far south of New Zealand.

Contributions from the Resident and Migrated cohorts were:

- Resident cohort:
 - for the far north 17.6 per 100,000 (95% CI: ± 3.0);
 - for the far south 66.6 per 100,000 (95% CI: ± 5.5).
- Migrated cohort:
 - for the far north 13.8 per 100,000 (95% CI: ± 2.6);
 - for the far south 33.5 per 100,000 (95% CI: ± 3.8).

These represent factors of 3.8 and 2.4, respectively, from the far north to the far south of New Zealand.

(Note that a factor of 2917/2005 must be included to allow for the missing cases if a direct comparison is to be made with prevalence rates from other studies.)

7.2.2 Prevalence–Latitude Gradient and Migration

External migration was controlled for by considering separate Migrated and Resident cohorts of cases; internal migration within New Zealand would be responsible for any differences between the prevalence–latitude gradients observed for different life periods for the Resident cohort.

The patterns of internal migration indicate that there was considerable migration both before and after onset of MS. Prior to onset of MS, 35% of the Resident cohort resided at the same latitude as at birth. Of the 65% who had some net migration within New Zealand, at onset of MS 74% resided within 2° latitude of their place of birth. Between onset of MS and census day 2006, 63% of the Resident cohort migrated from their location at onset of MS, with 83% of these by less than 2° latitude.

It was found that log least squares regression models (fits) to the prevalence–latitude data are more suitable than linear fits for the data in this study. (*p*-values given here are from ANOVA *F* tests of the regression models.)

For census day 2006, percentage increases in prevalence of MS per degree of latitude south are:

- Combined cohort 11.7% (95% CI: 11.0–12.4%; $p < 0.001$);
- Resident cohort: 14.0% (95% CI: 12.8–15.1%; $p < 0.001$);
- Migrated cohort: 8.4% (95% CI: 7.8–9.0%; $p < 0.001$).

The rates for the Resident cohort and the Migrated cohort are significantly different ($p = 0.004$). The rate obtained for the Combined cohort is equivalent to a weighted combination of the rates for the individual cohorts.

The rate of increase of prevalence of MS with latitude varies depending on:

- gender;
- MS phenotype;
- the history of residence locations of the cases considered;
- the time of residence being considered: location at census day 2006, or location during first five years of life (Age 0–4).

These factors are not independent; there is a complex interdependency between them and all must be taken as potential confounding covariables in any analysis. Percentage increases in prevalence of MS per degree of latitude south are given in Table 7.1.

Cohort	Sample	Period	Change in Prevalence per degree south	
			RR/SPMS (95% CI)	PPMS (95% CI)
Migrated	839	2006	8.1% (7.2–9.1%)	3.5% (0.0–7.2%)
Resident	1166	2006	15.2% (13.2–17.2%)	9.6% (6.5–12.8%)
Resident	1166	Age 0–4	14.5% (13.1–15.9%)	6.3% (3.6–9.2%)

(a) Female cases

Cohort	Sample	Period	Change in Prevalence per degree south	
			RR/SPMS (95% CI)	PPMS (95% CI)
Migrated	839	2006	7.6% (6.3–8.9%)	24.7% (22.7–26.8%)
Resident	1166	2006	9.5% (5.4–13.7%)	11.4% (7.1–15.8%)
Resident	1166	Age 0–4	10.0% (8.8–11.1%)	15.7% (14.6–16.7%)

(b) Male Cases

Table 7.1 Percentage rates of change of prevalence with latitude

Percentage increases in prevalence of MS per degree of latitude south by gender, MS phenotype, sample cohort and period of life. Period: 2006 is for census day 2006, reference population 2006 census; Age 0–4 is for the average residence location for the first five years of life, reference population 1956 census. Prevalence is age-standardised to the European standard with the appropriate reference population.

The rates of change of prevalence of MS with latitude given in Table 7.1 suggest a number of trends:

- Rates for female PPMS cases are lower than those for female RR/SPMS cases, for all cohort–life period combinations;
- Rates for male PPMS cases are higher than those for male RR/SPMS cases, for all cohort–life period combinations;
- The rate for male PPMS cases in the Migrated cohort (24.7%) is higher than all other rates; rates for all other Migrated gender–phenotype combinations are lower than all other rates.

The log regression fits for different gender–MS phenotype combinations, and for different cohorts and life periods, were compared using Anova tests at 95% CI.

Referring to rate ratios given in Table 7.1, the two most notable results are for the Resident cohort:

- For the Resident cohort and residence locations at Age 0–4, the rate ratios differ significantly for all four relevant comparisons:
 - The female and male RR/SPMS rates differ ($p = 0.047$);
 - The female and male PPMS rates differ ($p = 0.023$);
 - The female RR/SPMS and PPMS rates differ ($p = 0.043$);
 - The male RR/SPMS and PPMS rates differ ($p = 0.011$).
- For the Resident cohort and residence locations on census day 2006 none of the rate ratios are significantly different for the same comparisons.

Within the Migrated cohort, and comparing the Migrated and the Resident cohorts:

- For the Migrated cohort and residence locations on census day 2006, the rate ratios differ significantly for two of the four relevant comparisons:
 - The female and male PPMS rates differ ($p = 0.003$);
 - The male RR/SPMS and PPMS rates differ ($p < 0.001$).

Both of these differences are driven by the exceptionally high rate of change of prevalence with latitude exhibited by the Male PPMS cases within the Migrated

cohort. The rate of 24.7% per degree of latitude equates to a factor of ten (9.96; 95% CI: 8.39–11.82%) from the far north to the far south of New Zealand.

- For residence locations on census day 2006, and comparing rates for the Migrated cohort with those for the Resident cohort, the rate ratios differ significantly for two of the four relevant comparisons:
 - The female RR/SPMS rates differ ($p = 0.017$);
 - The male PPMS rates differ ($p = 0.029$).

The log regression fits for the prevalence–latitude data for census day 2006 are not significantly different from the fits for the Age 0–4 data for any of the gender–MS phenotype combinations. However, the internal migration which occurred from birth to census day 2006 was sufficient to change the significant differences in gender–MS phenotype behaviour observed for the Age 0–4 residence locations to non-significant differences in behaviour for census day 2006.

7.2.3 Month of Birth

Prevalence of MS depends upon month of birth, with gender and MS phenotype as covariables. Month of birth data for the Resident cohort was normalised by the average number of births per month for New Zealand averaged over 1980–1999. No analyses were undertaken to determine uncertainties or the significance of results. Due to low sample numbers, latitude was simply considered as north and south with a divide at 41°S. See Figure 5.27 for full details of the results.

- Female RR/SPMS cases show low numbers for February births in both the north and the south. They also show high numbers for November and December births in the north, and October to December births in the south.

Sample numbers are lower for male RR/SPMS and female PPMS cases. Patterns in most of the monthly data may be due to random variations. However, the following trends are apparent:

- Male RR/SPMS cases are low for May to July births in the north and high for June and July births in the south. The average for north plus south shows little or no variation in numbers;

- Female PPMS cases are low for March to May in the north and high for April to June births in the south;
- Both male RR/SPMS and female PPMS cases are low for February births in the south.

For male PPMS cases, there were only 1–5 cases per month for the north, and 0–7 cases per month for the south. Any visible patterns in the monthly data for male PPMS cases may be due to random variations. Allowing for the variations in the data, in the south there is a general trend over the yearly cycle for low numbers for April and May births, and high numbers for November to January births.

7.3 Implications of Results

Prior to this study, and the initial results of Taylor et al (2010b), none of the MS prevalence studies which looked at the prevalence–latitude gradient differentiated between gender and MS phenotype. Taylor et al established that the linear gradients varied by gender and MS phenotype; the later meta-analysis study by Simpson et al (2011) further established that the log gradients, which give the rate of change of prevalence with latitude, differ by gender. In addition to establishing that the rates of change of prevalence with latitude vary by both gender and MS phenotype, this study looks at the effect of lifetime migration on the prevalence–latitude gradients. It also examines whether latitude can be considered as a surrogate for average UV exposure. Implications of the findings will now be discussed, starting with UV exposure.

7.3.1 UV Exposure and Vitamin D

Although a multitude of other genetic and environmental factors are believed to be involved in risk of MS, vitamin D status is a factor which is often linked to latitude through UV exposure, with considerable evidence linking vitamin D deficiency, or insufficiency, to many auto-immune related diseases and disorders, including MS (Kampman and Steffensen 2010; Munger et al. 2006; Ponsonby et al. 2005; Van Belle et al. 2011). Other photochemical products which may be involved with risk of MS include urocanic acid. Biological pathways involved with urocanic acid, however, are just as complex than those of vitamin D. Therefore, discussions here will concentrate on links with vitamin D, with other factors referred to where required.

It has been established in this study that average monthly UV exposure levels vary linearly with latitude, and hence, latitude can be considered as a surrogate for average UV exposure within New Zealand. However, the photosynthesis of vitamin D from UV radiation is just one part of a complex biochemical equilibrium. At any given time, as well as depending on the incident UV radiation, this balance also depends on skin type (which affects the rate of photosynthesis), the concentrations of other reagents within the equilibrium process, and the current blood serum levels of vitamin D (vitamin D status). Vitamin D status, in turn, also involves the amount of body fat (which determines vitamin D storage) and a natural reduction through further metabolism, with a half-life of around two months. As such, the relationship between UV exposure and vitamin D status is not trivial. In addition, dietary intake of vitamin D will confound any relationship by affecting the vitamin D status directly. However, in the absence of sufficient dietary intake of vitamin D, it is accepted that low exposure to UV will lead to vitamin D deficiency, with a sufficient vitamin D status only arising from suitable exposure to UV. Therefore, it is likely that, in the absence of sufficient dietary intake of vitamin D, vitamin D status will be dependent on latitude, subject to modification by variations in actual UV exposure on an individual basis. A further complication is that, due to the complex biological pathways involved in the photosynthesis and storage of vitamin D, measures such as blood serum levels of vitamin D will lag measured environmental factors such as UV exposure within any annual cycle. These lags must be considered when analysing time dependent data such as month of birth prevalence rates in order to associate potential biological factors with periods of risk and of benefit during pregnancy. In addition, it is important to understand that the situation described is for vitamin D—the duration and nature of any lags are likely to vary for different photosynthetic agents.

As an example, consider female RR/SPMS cases which show high numbers of births from October to December in the south. The months for lowest UV exposure are May to July in New Zealand, however, the months for lowest maternal vitamin D status are likely to be July to September, assuming a lag of two months. If vitamin D deficiency during pregnancy is involved in determining risk of MS then the period of risk during foetal development will be roughly two months later than that suggested by UV exposure levels, that is, around the end of the second trimester and the beginning of the third, rather than the first half of the second trimester. Similarly, the high in the

number of female PPMS births in April and May doesn't appear to correlate with low UV exposure during pregnancy. However, if low maternal vitamin D status is the important factor then a July to September low in vitamin D status would actually correspond to the weeks following conception, the earliest stage of foetal development.

The large study by Willer et al (2004) looked at month of birth effects by combining data from several countries to give data for over 42 thousand cases. The countries involved were Canada, Denmark, Sweden and the UK, all at higher latitudes than New Zealand. As such, results are more likely to compare with the month of birth results obtained here for the south half of New Zealand, rather than the whole or the north. However, within the Willer et al study, there was no breakdown of results by gender or MS phenotype. A peak in risk was found for May births, which does coincide with the broad increase in female RR/SPMS births around November in this study, though the drop in risk observed for November in the Willer et al study does not appear to have an equivalent drop in May births in this study. Willer et al also state in their discussion that "the abrupt change in risk by month suggests a threshold effect for both increased and decreased risk, something that is not easily explained". However, such a threshold effect can be easily explained if there are opposing effects in different groups of cases (for example, different gender-MS phenotype) which are then combined. Similar month of birth results by Staples et al (2010) and Salzer et al (2010) also took no account of gender or phenotype.

A further consideration for vitamin D and UV exposure relates to the period of life under study. Here, sun exposure data was collected for all periods from birth to onset of MS. If data relating to sun exposure had simply been collected for the period of life leading up to census day 2006, that is, after onset of MS, then the data is likely to reflect other health factors due to the presence of MS rather than residence location. Van der Mei et al (2007a) established that people with MS in Tasmania had lower vitamin D levels than controls. However, this was also strongly associated with the level of disability of the cases; cases with higher disability also tended to have lower exposure to sun, probably leading to a lower vitamin D status. In order to study the link between vitamin D status and the risk of MS, a study such as that by Munger et al (2006) is required, where the vitamin D status prior to onset of MS is examined.

Although the NZMSPS project collected data for time spent outdoors by participants during each period of their life up to onset of MS, the request was for hours per week, without regard to time of day. Since the time of day of any exposure to sunlight seriously affects the UV dosage received, methods to incorporate this data into the study were not pursued. This will be discussed further in section 7.4 along with other limitations of this study.

7.3.2 Prevalence of MS, Latitude Gradients, and Migration

There are significant variations in the prevalence of MS with latitude depending upon gender, MS phenotype and the cohort of cases under study. Further, considering the Resident cohort at Age 0–4, high latitude at birth is:

- a high risk factor for RR/SPMS for females;
- a low risk factor for RR/SPMS for males;
- a low risk factor for PPMS for females;
- a high risk factor for PPMS for males.

It is well established that there is a gradient of the prevalence of MS with latitude, other factors being equal (Ebers and Sadovnick 1993; Risco et al. 2011; Simpson et al. 2011; Wallin et al. 2004). However, this is the first study to establish that the rate of change of prevalence of MS by latitude differs by both gender and MS phenotype. That is, that latitude (as a surrogate) acts differently according to gender and MS phenotype. As a consequence, gender and MS phenotype must both be allowed for in any studies which aim to establish relationships between the prevalence of MS and any factors which may depend on latitude, such as UV exposure and vitamin D status.

The study in eastern Australia by Taylor et al (2010a) established similar gradient effects when looking at the incidence of first CNS demyelinating events (FDEs), a precursor to MS. The FDE incidence varied by gender, FDE subtype and latitude, with overall incidence rates increasing by 9.55% per degree of latitude. Any correlation between the effect of latitude on the incidence of FDE and the effect of latitude on the prevalence of MS, whilst taking gender, FDE subtype and MS phenotype into consideration, could help towards an understanding of the aetiology of MS.

Just as the effect of gender and MS phenotype must be considered when looking at latitude effects on MS, so the effect of latitude must not be ignored when looking at other factors. For example, month of birth data for male RR/SPMS cases show opposing winter trends when considering the north and the south. An average for the north plus the south only shows variations which could easily be random, thus hiding sizeable effects which are only highlighted in the individual north and south cohorts.

Migration is a significant confounding factor. Most studies which have looked at migration have concentrated on the effect of age at migration in conjunction with the origin and target residence locations (Alter et al. 1978; Elian et al. 1990; McLeod et al. 2011; Wallin et al. 2009). However, Cabre et al (2005) also considered the effect of migrants returning to the French West Indies from countries such as France, where the prevalence of MS is much higher. For that study, where 33% of the population were migrants, the prevalence of MS was higher in the migrants than in the non-migrants by a factor of 2.2. In this study, where migrants are migrating to an area of high prevalence, the migrant group has the lower rate of prevalence and the resident group has the higher rate, by a factor of 2.0 for the far south of New Zealand. In both studies, the presence of migrants within the population affects the prevalence of MS within that population. If the prevalence of MS is required in order to establish the burden of disease for that area, then migration effects need not be considered. If, however, factors are being studied in order to understand the aetiology of MS then migration effects must be taken into account when determining the prevalence of MS.

In addition to allowing for external migration, care must be taken when controlling for migration if effects which could vary with migration are to be analysed. In this study, there was a balance between controlling for external migration, whilst allowing internal migration. By looking at the residence locations of the Resident MS cases at ages 0–4, it was possible to see how the prevalence–latitude gradients varied with gender and MS phenotype without the confounding effects of external migration, and prior to the confounding effects of internal migration. The lack of any significant differences between the prevalence–latitude gradients for the same cohort for census day 2006, illustrates the confounding effect of even internal migration. Prevalence–latitude studies which do not control for internal migration may be allowing migration to confound their results.

7.4 Strengths and Limitations of the Study

This study was based on data collected as part of the NZMSPS project, which was able to identify an estimated 96% of the cases of MS within New Zealand at the time of the project. Although 31% of those cases were unable or unwilling to provide further data for this study, there were 2005 cases for which sufficient data was available, with 1166 of these being resident in New Zealand from birth to onset of MS. This enabled a variety of analyses to be undertaken which would not have been possible with smaller numbers. For example, the month of birth data could be analysed by gender, MS phenotype, and a measure of latitude of residence at ages 0–4. The results obtained suggest that other studies where data is not stratified by these factors, may be missing key pointers to how these factors interact in determining the prevalence of MS. An understanding of these interactions is, in turn, key to unravelling the aetiology of MS.

The number of cases involved in the study also enabled the prevalence–latitude gradient within New Zealand to be analysed in much more depth than it had been before. In addition to being able to analyse the gradient for the different gender–MS phenotype combinations, it was possible to examine how the results varied with the data aggregation methods used, and to select the method which provided results with the tightest confidence bands. In reducing the uncertainties, this also gave a greater number of significant results within the comparisons which were made.

The study is based on the cases of MS resident in New Zealand on census day 2006. As such, any MS cases who were born in New Zealand and then either emigrated or died before census day 2006 would not have been included in the Age 0–4 analyses. Further selection bias may have been introduced if the 31% missing cases differed from the cases which remained in the Resident and Migrated cohorts. One known difference is ethnicity since a disproportionate number of Māori cases were in this group, thus preventing any further analysis of ethnicity.

There were two limitations within this study which relate to the sunlight exposure data collected as part of the NZMSPS project. The data collected was only specified as hours per week spent outdoors, with no regard to the time of day, even though UV radiation exposure from sunlight varies considerably by time of day. Although this limitation in itself would not have been sufficient to prevent the data being analysed as a measure of average UV exposure, the lack of control data prevented any analyses comparable

with those used to establish the prevalence–latitude gradients. This is discussed further in the next section.

7.5 Further Research

There are several areas in which this study needs to be extended. First, the data for the years prior to onset of MS needs to be analysed to give prevalence–latitude gradients for this period. The currently missing pre-requisite for this—the census boundary data for a suitable census, for example 1986—is easily achievable, given further resources. Adding these gradients to those already obtained for the Age 0–4 period and for census day 2006, will aid an understanding of how latitude (as a surrogate) may act as a risk factor prior to onset of MS.

The migrated cohort of cases must be analysed further. Many who are born in New Zealand spend time in Australia before returning to New Zealand. Stratification of this cohort according to certain, controlled, migration patterns should point to which factor, or factors, appear to remove much of the latitude effect from the prevalence–latitude gradients obtained for all but the male PPMS cases. In addition, the male PPMS cases should be examined further to ascertain risk factors which may have given rise to the exceptionally high rate of change of prevalence obtained for this group.

If suitable control data is obtained for time spent outdoors, and hence, exposure to sunlight, then the sunlight exposure data collected for this study could be analysed. The study by Lucas et al (2011b) looking at FDEs in Australia, collected sun exposure data from both cases and controls, as well as vitamin D serum levels. The study was able to establish that both sun exposure and vitamin D levels are independently associated with the risk of FDE. Suitable control data for this study would enable similar sun exposure analyses for the prevalence of MS to be extended into the higher latitudes of southern New Zealand where differences between effects should be more noticeable.

Although the prevalence–latitude gradients already obtained in this study are quite conclusive, there may be other risk factors for MS which depend on latitude in some way other than through UV exposure. One such factor is Scottish ancestry: the study of Miller et al (1990) highlighted the preferential settlement of Scots in the south of New Zealand rather than the north. The prevalence of MS in Scotland and northern parts of

Ireland tends to be greater than England and southern Ireland, with variations which are unlikely to be due to latitude alone (McGuigan et al. 2004; Pugliatti et al. 2006). This may be due to a higher frequency of the HLA DR15 haplotype in the Scottish and northern Irish (McGuigan et al. 2004; Swigler and Compston 1986). Since alleles in this haplotype have been proposed as risk factors for MS in other populations (Harbo et al. 2007; Schmidt et al. 2007; Spurkland et al. 1997) then having Scottish ancestry could be a risk factor, which might vary with latitude within New Zealand. A more complete analysis of HLA allele frequencies, would be a useful addition to any further analyses.

7.6 Conclusions

This study has confirmed that there is a latitudinal gradient in the prevalence of MS within New Zealand. New findings establish that the rate of change of prevalence with latitude varies depending upon gender and MS phenotype within residents of New Zealand who have not migrated into or out of New Zealand. The rates are significantly different when looking at location of residence at age 0–4, and not significantly different when taking prevalence on census day 2006. For census day 2006, the rates of change of prevalence with latitude for the cohort who had migrated into or out of New Zealand were different from the rates for those who had not migrated; the differences were significant for female RR/SPMS and male PPMS cases. These findings indicate that the genetic and environmental factors underlying the prevalence of MS act differently by gender and MS phenotype and that internal migration within New Zealand, in addition to external migration, is a serious confounding factor when looking at latitudinal data. The rates of change of prevalence with latitude for the period of life prior to onset of MS should be determined for those who remained within New Zealand in order to clarify how migration pre onset of MS and post onset of MS affects the rates obtained.

The prevalence of MS within New Zealand also varies by month of birth. Trends observed indicate that different periods of pregnancy are important for the different gender–MS phenotype combinations. These trends also vary by latitude of residence.

Future studies which look at prevalence of MS with an aim to understanding the aetiology of MS must allow for gender, MS phenotype, and location of residence in conjunction with migration, in addition to any other factors of concern.

Appendix A The NZMSPS Questionnaire

The NZMSPS questionnaire was a 20 page document covering four sections. The first three sections were: the main questionnaire; the residence calendar; and the job calendar. The remaining section contained two consent forms which would be held separately from the rest of the questionnaire to ensure the confidentiality of the case. The first form asked for consent to enable the study centre to contact the case as part of the diagnostic process; the second form asked for consent to enable the study centre to contact the case either as a precursor to future research, or in the event that a national register of people with MS was set up.

The pages from the questionnaire are reproduced (at reduced size) in the following pages of this appendix.

□ □ □ □ □ □ □ □ □ □



New Zealand Multiple Sclerosis Prevalence Study Questionnaire.

Thank you for taking the time to answer this questionnaire about MS in New Zealand. The first section of the questionnaire is about you. Many of the questions are standard questions about education, marital status, housing and income, derived from the New Zealand census. We need this information to find out how people with MS compare with the New Zealand population as a whole.

Some of the questions ask for personal information. We can assure you that this information will be kept completely confidential. You will not be able to be identified in any report or publication of the results of this study. If you do not wish to answer a particular question please put a line through it rather than leaving it blank, to let us know that you do not wish to answer that question. Please remember however, that the more complete the information we receive, the better the information we will be able to produce.

A number of questions ask about your heritage. We are interested in knowing about your ancestry because some ethnic groups appear to be protected from developing MS. We are interested in knowing if this is true within the New Zealand population.

A number of environmental factors have also been identified as possibly increasing the likelihood of developing MS. For this reason, we would like to know about the places you have lived and worked. The calendars in the second section of the questionnaire are designed to find out about this. You may wish to ask your family for help with this section, especially your earliest addresses and outdoor activities. An example of how to fill out each calendar is given at the top of each page.

If you require help with any part of the questionnaire please do not hesitate to contact the study team on phone number: **0800 MS STUDY (0800 677 8839)**, or e-mail: **msstudy@chmeds.ac.nz**. We would be happy to assist you.

□ □ □ □ □ □ □ □ □ □

Section 1: Who you are

1. What age group are you in?
(Please tick the appropriate box)

- Less than 20 years
- 20-24
- 25-29
- 30-34
- 35-39
- 40-44
- 45-49
- 50-54
- 55-59
- 60-64
- 65-69
- 70-74
- 75 and older

For office use

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13

2. a. Where did you live at the time of the 2006 census (March 2006)?
(Please print the suburb, city or town, and region below)

Suburb _____

City or town _____

Region _____

□ □ □
14 15 16



b. Where were you born ? *(Please print the suburb, city or town, and region if born in NZ or town/city and country if overseas)*

Suburb _____

City or town _____

Region (or country) _____

3. a. What year were you first diagnosed with MS? _____

 17 18 19 20

b. Where were you living? *(Please print the suburb, city or town, and region if born in NZ or town/city and country if overseas)*

Suburb _____

City or town _____

Region (or country) _____

4. a. What year did you first have symptoms of MS? _____

 21 22 23 24

b. Where were you living? *(Please print the suburb, city or town, and region if born in NZ or town/city and country if overseas)*

Suburb _____

City or town _____

Region (or country) _____



Many of the following questions relating to ethnicity, occupation and living situations are identical to those in the 2006 census enabling us to make comparisons between people with MS and the general population of NZ.

5. Which ethnic group do you belong to?
(Please tick the box or boxes which apply to you)

- | | |
|---|---|
| <input type="checkbox"/> 1. NZ European | <input type="checkbox"/> 2. Maori |
| <input type="checkbox"/> 3. Samoan | <input type="checkbox"/> 4. Cook Island Maori |
| <input type="checkbox"/> 5. Tongan | <input type="checkbox"/> 6. Niuean |
| <input type="checkbox"/> 7. Chinese | <input type="checkbox"/> 8. Indian |
| <input type="checkbox"/> 9. Other (such as Dutch, Japanese, Tokelauan). | |

Please state: _____

6. What is your ancestral group? This refers to your ancestral origins (heritage) and may be different from your ethnic group.
(Please tick as many boxes as you need).

- | | |
|--|-----------------------------|
| <input type="checkbox"/> 1. European | <input type="checkbox"/> 25 |
| <input type="checkbox"/> 2. Maori (Iwi/Hapu: _____) | <input type="checkbox"/> 26 |
| <input type="checkbox"/> 3. Samoan | <input type="checkbox"/> 27 |
| <input type="checkbox"/> 4. Cook Island Maori | <input type="checkbox"/> 28 |
| <input type="checkbox"/> 5. Other | <input type="checkbox"/> 29 |
| Please state: _____ | <input type="checkbox"/> 30 |
| | <input type="checkbox"/> 31 |
| | <input type="checkbox"/> 32 |
| | <input type="checkbox"/> 33 |
| | <input type="checkbox"/> 34 |
| | <input type="checkbox"/> 35 |
| | <input type="checkbox"/> 36 |
| | <input type="checkbox"/> 37 |
| | <input type="checkbox"/> 38 |

NZMS Prevalence Study □ □ □ □ □ □ □ □ □ □

7. As best you can, please give your grandparents' ancestry
(Please tick the box or boxes which apply to each grandparent)

		MF
		<input type="checkbox"/> 39
		<input type="checkbox"/> 40
Maternal Grandparents (your mother's parents)		<input type="checkbox"/> 41
		<input type="checkbox"/> 42
		<input type="checkbox"/> 43
		MM
		<input type="checkbox"/> 44
		<input type="checkbox"/> 45
		<input type="checkbox"/> 46
		<input type="checkbox"/> 47
		<input type="checkbox"/> 48
		FF
Paternal Grandparents (your father's parents)		<input type="checkbox"/> 49
		<input type="checkbox"/> 50
		<input type="checkbox"/> 51
		<input type="checkbox"/> 52
		<input type="checkbox"/> 53
		FM
		<input type="checkbox"/> 54
		<input type="checkbox"/> 55
		<input type="checkbox"/> 56
		<input type="checkbox"/> 57
		<input type="checkbox"/> 58

Questionnaire 2006 03 08 5

NZMS Prevalence Study □ □ □ □ □ □ □ □ □ □

8. What is your highest secondary school qualification?

1. None

2. NZ School Certificate in one or more subjects, or National Certificate level 1 or NCEA Level 1

3. NZ Sixth Form Certificate in one or more subjects, or National Certificate level 2 or NZ UE before 1986 in one or more subjects or NCEA Level 2

4. NZ Higher School Certificate, or Higher Leaving Certificate or NZ University Bursary / Scholarship or National Certificate Level 3 or NCEA Level 3 or NZ Scholarship level 4

5. Other secondary school qualification **gained in NZ.**
Print what it is: _____

or

6. Other secondary school qualification **gained overseas**
Print what it is: _____ 59

9. Apart from secondary school qualifications, do you have another completed qualification? *DON'T count qualifications that take less than 3 months of full-time study to get*

Yes No

Please go to question 10 **Please go to question 11** 60

10. Print your highest qualification, and the main subject, for example:
Qualification *TRADE CERTIFICATE*
Subject *ELECTRICAL ENGINEERING*

Qualification (and level, if applicable)

Subject _____ 61

Questionnaire 2006 03 08 6



11. In the last 7 days, which of these did you do?
(Tick as many boxes as you need to answer this question)

- I worked for pay, profit or income **Please go to question 12** 62
- I worked in a family business or family farm **without pay** **Please go to question 12** 63
- None of these. **Please go to question 14** 64

12. In that job, what was your occupation, for example:
*PRIMARY SCHOOL TEACHER, CLOTHING MACHINIST,
MOTEL MANAGER, RECEPTIONIST?*

_____ 65

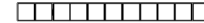
13. How many hours, to the nearest hour,
in all your jobs (for profit or unpaid in
a family business/farm) do you usually
work each week? _____ hours

66 67 68

14. Has your job (occupation or hours of work) ever changed as a result of MS?

Yes **If yes, please explain why:**

No **If no, please go to question 16** 69



15. Please specify how your job has changed for the most recent time as a result
of MS (Tick as many boxes as you need)

- Changed jobs 70
- Ceased working 71
- Reduced number of working hours worked per week 72
- Increased number of hours worked per week 73
- Commenced sickness benefit 74
- Commenced invalids benefit 75
- Became student 76
- Other (please specify _____) 77

16. What was the occupation you were **originally** trained for?
For example: *PRIMARY SCHOOL TEACHER, CLOTHING
MACHINIST, MOTEL MANAGER, RECEPTIONIST?*

_____ 78

NZMS Prevalence Study		[] [] [] [] [] [] [] [] [] [] [] []									
17.	Tick as many boxes as you need to show all the ways you yourself got income in the 12 months ending today (DON'T count loans because they are not income):										
<input type="checkbox"/>	Wages, salary, commissions, bonuses, etc, paid by my employer	<input type="checkbox"/>	79								
<input type="checkbox"/>	Self-employment, or business I own and work in	<input type="checkbox"/>	80								
<input type="checkbox"/>	Interest, dividends, rent, other investments	<input type="checkbox"/>	81								
<input type="checkbox"/>	Regular payments from ACC or a private work accident insurer	<input type="checkbox"/>	82								
<input type="checkbox"/>	New Zealand Superannuation or Veterans Pension	<input type="checkbox"/>	83								
<input type="checkbox"/>	Other superannuation, pensions, or annuities (other than NZ Superannuation, Veterans Pension or war pensions)	<input type="checkbox"/>	84								
<input type="checkbox"/>	Unemployment Benefit	<input type="checkbox"/>	85								
<input type="checkbox"/>	Sickness Benefit	<input type="checkbox"/>	86								
<input type="checkbox"/>	Domestic purposes benefit	<input type="checkbox"/>	87								
<input type="checkbox"/>	Invalids Benefit	<input type="checkbox"/>	88								
<input type="checkbox"/>	Student Allowance	<input type="checkbox"/>	89								
<input type="checkbox"/>	Other government benefits, government income support payments, war pensions, or paid parental leave	<input type="checkbox"/>	90								
<input type="checkbox"/>	Other sources of income, counting support payments from people who do not live in my household	<input type="checkbox"/>	91								
	or										
<input type="checkbox"/>	No source of income during that time	<input type="checkbox"/>	92								
Questionnaire 2006 03 08		9									

NZMS Prevalence Study		[] [] [] [] [] [] [] [] [] [] [] []									
18.	From all the sources of income you marked in question 17, what was the total income that you yourself got before tax or anything was taken out of it, in the 12 months ending 31 March 2006?										
<input type="checkbox"/>	1. Loss										
<input type="checkbox"/>	2. Zero income										
<input type="checkbox"/>	3. \$1 - \$5,000										
<input type="checkbox"/>	4. \$5,001 - \$10,000										
<input type="checkbox"/>	5. \$10,001 - \$15,000										
<input type="checkbox"/>	6. \$15,001 - \$20,000										
<input type="checkbox"/>	7. \$20,001 - \$25,000										
<input type="checkbox"/>	8. \$25,001 - \$30,000										
<input type="checkbox"/>	9. \$30,001 - \$35,000										
<input type="checkbox"/>	10. \$35,001 - \$40,000										
<input type="checkbox"/>	11. \$40,001 - \$50,000										
<input type="checkbox"/>	12. \$50,001 - \$70,000										
<input type="checkbox"/>	13. \$70,001 - \$100,000										
<input type="checkbox"/>	14. \$100,001 or more										
		[] []									
		93 94									
Questionnaire 2006 03 08		10									



19. Which of these statements is true about your legal marital/civil union status? (If you have had more than one legal marriage/civil union, answer for your most recent).

- 1. I have never been legally married and I have never been legally joined in a civil union
- 2. I am divorced or my marriage has been dissolved
- 3. I am a widow / widower / bereaved civil union partner
- 4. I am permanently separated from my legal husband / wife / civil union partner
- 5. I am legally married
- 6. I am legally joined in a civil union

95

20. Where do you live at present? (Please tick one box)

- 1. A house or flat
- 2. Boarding house
- 3. Hospital
- 4. Rest home
- 5. Other (please state: _____)

96



21. Where did you live 5 years ago? (Please tick one box)

- 1. A house or flat
- 2. Boarding house
- 3. Hospital
- 4. Rest home
- 5. Other (please state: _____)

97

22. At present, do you receive services from any of the following? (Please tick as many boxes as you need)

- District nurse
- Physiotherapist
- Home care
- Meals on wheels
- Multiple Sclerosis Society
- Other (please state: _____)
- None

98

99

100

101

102

103

104

NZMS Prevalence Study □ □ □ □ □ □ □ □ □ □

23. Does anyone in your immediate family (parents, siblings, half-siblings) have, or has anyone had multiple sclerosis?

(Please tick as many boxes as you need. If you have more than one brother or sister with MS please write the number of brothers or sisters with MS in the space provided below)

No 105

Mother 106

Father 107

Brother(s) Number of brothers with MS _____ 108

Sister(s) Number of sisters with MS _____ 109

Half-Brother(s) Number of half-brothers with MS _____ 110

Half-Sister(s) Number of half-sisters with MS _____ 111

Thank you, this concludes section 1

Questionnaire 2006 03 08 13

NZMS Prevalence Study □ □ □ □ □ □ □ □ □ □

Section 2: Residence Calendar

There has been some suggestion that the likelihood of developing MS may be related to where a person has lived in their lifetime. For this reason, we would like to know more about the places where you were brought up and have lived. The calendar on the following pages asks you to record these locations. By combining the data from all the participants in this study, we hope to identify patterns that will help us understand MS better.

- The calendar asks you to describe (in chronological order) where you have lived, and for how long you stayed at each location. An example of how to fill out the calendar is included at the top of each page.
- Please include all places you have lived, **from the start of your mother's pregnancy with you (your conception) to the date at which you were first diagnosed with MS**. Only include places where you have lived for more than 6 weeks. Please make every effort to ensure completeness, especially in your early life. If possible we would encourage you to consult other family members to help fill in details.
- Please start a new row for each time you moved, and continue on a separate page if you need more rows.

From & To section:
Please include **both** the month and the year. If you can't recall the exact month, please estimate. Don't leave blanks.

Where you have lived section:
For urban areas, fill in the **suburb and city**. For rural areas, fill in the **closest town and district**. If the address is **not in NZ** just fill in the **town and the country**. Please don't use PO Box numbers.

Sunlight exposure section:
For each place you have lived we would like to estimate how much sunlight you have been exposed to. This has probably changed at different times during your life. We therefore ask that for each place you have lived, separately for summer and winter, you tick the box that best estimates how many hours a week, on average you would normally have spent in the sun.

For each period of your life we are trying to capture an overall measure of how much time you spent in the sun. If you are unsure about your average sun exposure during any period, it may help to think about where you were and what you were doing at that time. For example, when you were at school maybe you played outside after school or at lunchtime; if you were working, did your job involve being indoors or outdoors a lot. During weekends or holiday periods did you spend more time indoors or outdoors.

You may wish to check with your family, particularly parents or siblings, about your sunlight exposure for ages before 6 as these are often difficult to recall.

If you are truly unsure and cannot find the information out, put a line through the sun exposure questions. **DO NOT LEAVE BLANK.**

Sunburn section:
For each period (each row), how many times were you sunburnt, **where the pain lasted two or more days?**

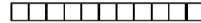
Questionnaire 2006 03 08 14

NZMS Prevalence Study																			
Office use only Period	From (Month and year)	To (Month and year)	Where you've lived Suburb and city, or nearest town and district		Sunlight Exposure Average hours per week spent in the sun Place tick ✓ in appropriate box										Sunburn Number of times sunburnt Place tick ✓ in appropriate box				
	1 Until diagnosis	2 Until diagnosis	3 Suburb / Town	4 City / District (or country)	Winter					Summer					15	16	17	18	19
						5 <5	6 5-9	7 10-19	8 20-29	9 30+	10 <5	11 5-9	12 10-19	13 20-29	14 30+	0	1	2-5	6-10
example	6 / 1962	12 / 1975	Highbury	Wellington															
R1	Conception	_ / _																	
R2	_ / _	_ / _																	
R3	_ / _	_ / _																	
R4	_ / _	_ / _																	
R5	_ / _	_ / _																	
R6	_ / _	_ / _																	
R7	_ / _	_ / _																	
R8	_ / _	_ / _																	
R9	_ / _	_ / _																	
R10	_ / _	_ / _																	
R11	_ / _	_ / _																	
R12	_ / _	_ / _																	
R13	_ / _	_ / _																	
R14	_ / _	_ / _																	
R15	_ / _	_ / _																	

Questionnaire 2006 03 08 15

NZMS Prevalence Study																			
Office use only Period	From (Month and year)	To (Month and year)	Where you've lived Suburb and city, or nearest town and district		Sunlight Exposure Average hours per week spent in the sun Place tick ✓ in appropriate box										Sunburn Number of times sunburnt Place tick ✓ in appropriate box				
	1 Until diagnosis	2 Until diagnosis	3 Suburb / Town	4 City / District (or country)	Winter					Summer					15	16	17	18	19
						5 <5	6 5-9	7 10-19	8 20-29	9 30+	10 <5	11 5-9	12 10-19	13 20-29	14 30+	0	1	2-5	6-10
example	6 / 1962	12 / 1975	Highbury	Wellington															
R16	_ / _	_ / _																	
R17	_ / _	_ / _																	
R18	_ / _	_ / _																	
R19	_ / _	_ / _																	
R20	_ / _	_ / _																	
R21	_ / _	_ / _																	
R22	_ / _	_ / _																	
R23	_ / _	_ / _																	
R24	_ / _	_ / _																	
R25	_ / _	_ / _																	
R26	_ / _	_ / _																	
R27	_ / _	_ / _																	
R28	_ / _	_ / _																	
R29	_ / _	_ / _																	
R30	_ / _	_ / _																	

Questionnaire 2006 03 08 16



Section 3: Job Calendar

For all jobs where you have worked for more than 6 weeks up until the time you were diagnosed with MS, please could you:

- List the dates you worked in that job
- Tell us what the job was, and
- Indicate whether it was largely an indoor or outdoor job.

Office use only Period	From (Month and year)	To (Month and year)	Occupation	Indoor or Outdoor Job?	
	1	2		3	4
	Until diagnosis	Until diagnosis		Largely Indoor	Largely Outdoor
example	6 / 1962	12 / 1975	cleaner	✓	
J1	__/__/__	__/__/__			
J2	__/__/__	__/__/__			
J3	__/__/__	__/__/__			
J4	__/__/__	__/__/__			
J5	__/__/__	__/__/__			
J6	__/__/__	__/__/__			
J7	__/__/__	__/__/__			
J8	__/__/__	__/__/__			
J9	__/__/__	__/__/__			
J10	__/__/__	__/__/__			
J11	__/__/__	__/__/__			
J12	__/__/__	__/__/__			
J13	__/__/__	__/__/__			
J14	__/__/__	__/__/__			
J15	__/__/__	__/__/__			



Confirmation of the diagnosis of MS

As the diagnosis of MS is sometimes difficult some participants may need to be seen in a research clinic in their area. At that clinic you will be seen by a study doctor. This clinic visit will involve a neurological examination that is not painful or intrusive. No additional tests will be performed. Some people will not need to attend a clinic, but we may need further information, which we could obtain by telephone.

If you would be prepared to be contacted please write your contact details below. This information will be kept separate from the rest of the questionnaire to maintain your confidentiality.

Name: _____

Preferred Name: _____

Address: _____

Telephone: _____

E-mail: _____

Preferred Means of contact: _____

Preferred day/time of contact: _____

Specialists Name: _____

Name of GP: _____

PTO.



MS Research and Register

We would like to contact people with MS to find out if they would be prepared to take part in future research into MS and/or for their names to be included on a confidential national register of people with MS.

Would you be prepared to take part in future research into MS?

Yes No

Would you be prepared to be contacted if a confidential national register of people with MS is established?

Yes No

(Your contact details will be used only for the purpose you agreed to.)

Finally, we would be very interested in any thoughts or opinions you may have as to the cause of your MS (If you have any thoughts or opinions about this, please write them below)

_____ PTO



If you completed this form on behalf of someone with MS please tick box:

THANK YOU FOR ANSWERING THIS QUESTIONNAIRE

229

[This page intentionally left blank.]

Appendix B 1956 Census Data

The methods used to construct the 1956 New Zealand Census boundaries (described in section 4.2.3) are outlined in section B.1. Section B.2 provides several tables from the 1956 New Zealand population census publication which were useful in determining the population distributions.

B.1 Constructing the 1956 Census Boundaries

Two of the 1956 census maps (Figure 4.3 and Figure 4.4) were used as backdrops to the 2006 census meshblock areas. In situations where the 1956 census boundaries differed markedly from current meshblock boundaries and cut across them, the 2006 meshblocks were split, using the map image and adjacent boundaries as a guide. For example, Figure B.1(a) illustrates Tuapeka County, comparing the 2006 meshblocks boundaries and the map image of the 1956 boundaries. For almost all of the county the boundaries are in close agreement. However, the shaded meshblock, '3006000', had to be split to match the 1956 boundaries. For the Boroughs, Cities and Town Districts, it was often possible to use the names of current AUs as a guide along with Table 14 of Volume I of the census publication which lists Counties with their Interior Boroughs and Town Districts (see section 4.2.3.2). For example, the 1956 area Lawrence Borough was given an area of 2.49 km² in the census publication and the six 2006 meshblocks which comprise the Lawrence Area Unit have an area of 2.47 km². These are shown in Figure B.1(b). Given the close agreement of the two area sizes and the significantly larger adjacent meshblocks, those six meshblocks were taken as being a reasonable representation of the 1956 statistical area. Period maps were also used to ensure that the correct areas were chosen during this process. Various digital maps were downloaded from online sources and used within the advertised terms and conditions (Auckland City Libraries; Christchurch City Libraries; NZMCH). The Te Ara online version of the 1966 Encyclopedia of New Zealand (McLintock 1966) was a rich source of information and many period maps came from this resource. For some County statistical areas the census maps were not clear due to the resolution of the printed maps, as, for example, in the Auckland City, Wellington and Christchurch areas. In these cases, the process described above for the Boroughs, Cities and Town Districts was followed, as illustrated in Figure B.2 which shows districts in the south of Auckland City.

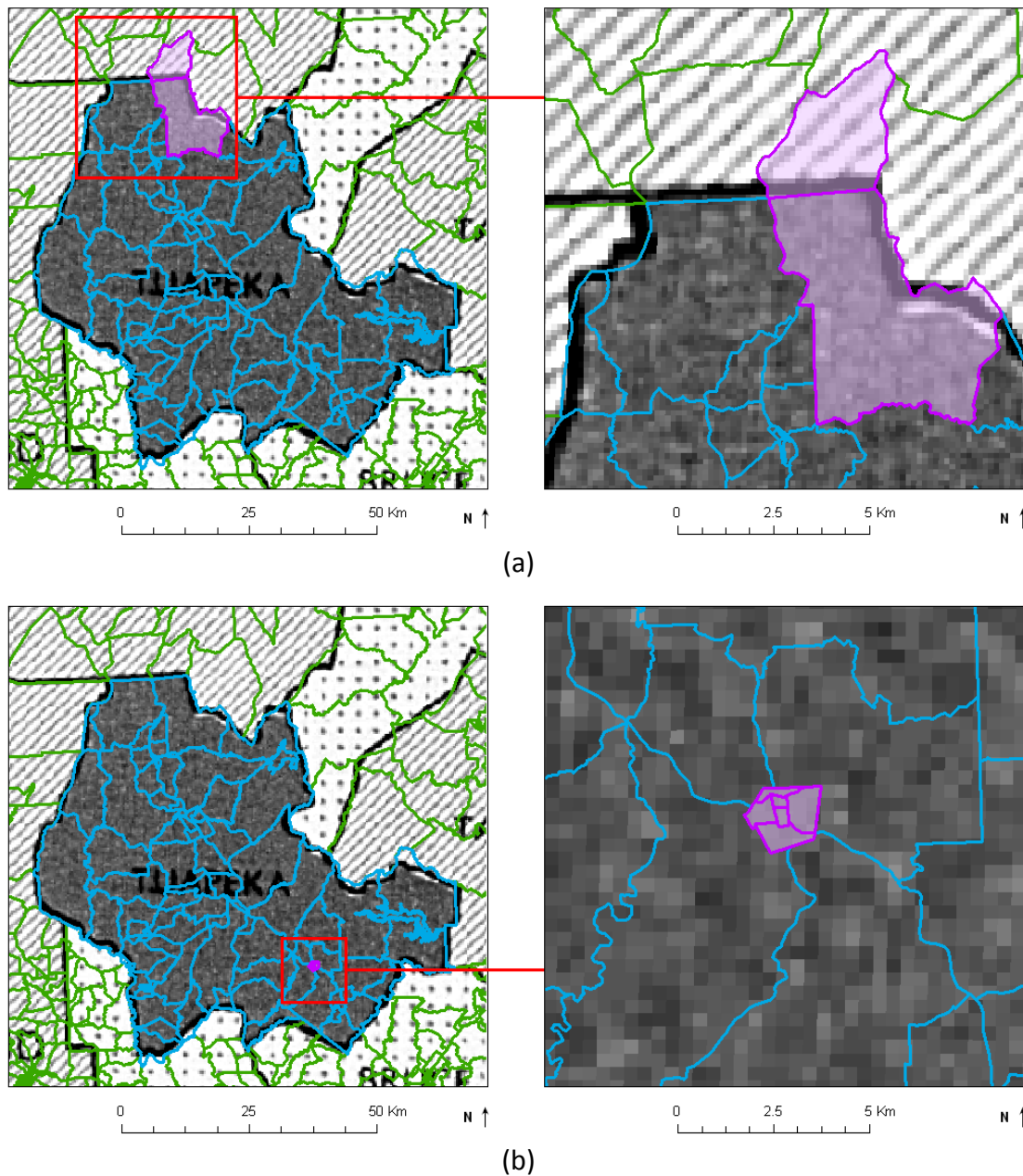


Figure B.1 Constructing the 1956 census areas from 2006 meshblocks

Tuapeka County, South Island, showing the generally good alignment of the 2006 meshblocks (blue) and the 1956 area boundaries (grey shaded raster backdrop).

(a) 2006 meshblock '3006000' (magenta) spans the boundary between the 1956 Tuapeka and Vincent Counties so was split for the 1956 areas. The constructed boundary continued the boundary between meshblocks to the west, meeting the eastern meshblock at its most westerly point. **(b)** The six 2006 meshblock areas which comprise the Lawrence Area Unit, with a combined area of 2.47 km² (magenta). These were taken to represent the 1956 Lawrence Borough statistical area which had an area of 2.49 km², the rasters were taken from the 1956 New Zealand population census publication (NZDoS 1959) and the meshblock data was downloaded from (StatsNZ n.d.).

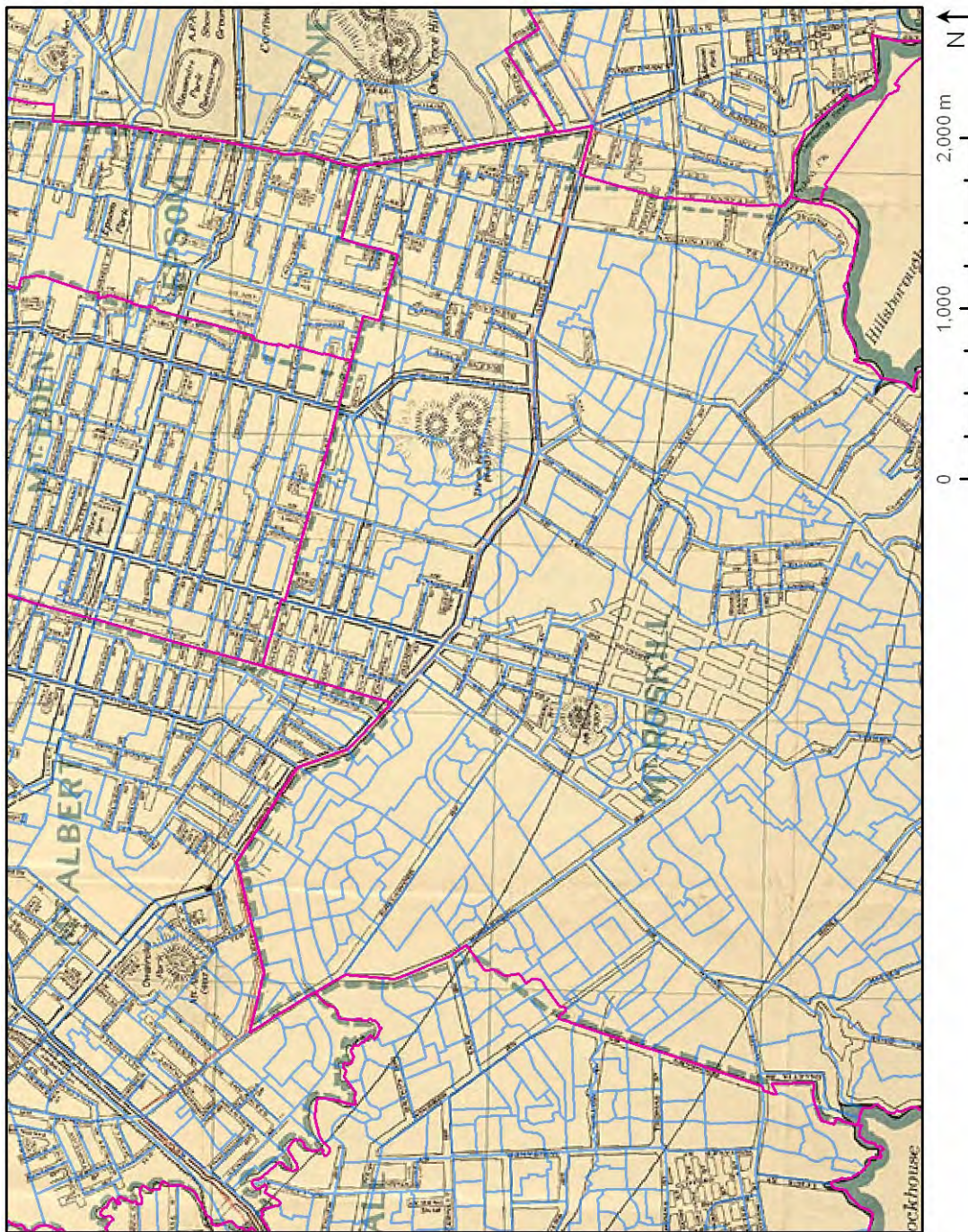


Figure B.2 Constructing the 1956 census areas for Auckland

Part of the south of Auckland City showing the generally good alignment of the 2006 meshblocks (blue) and the 1956 area boundaries (dashed heavy grey line in the raster backdrop). Magenta lines highlight the 1956 areas. Four areas can be seen marked: Epsom, Mt Albert, Mt Eden and Mt Roskill. Areas not fully marked here: Avondale (west), One Tree Hill (east) and Onehunga (south east). The areas bounded by the solid heavy grey line (south) are part of Manukau Harbour. The raster image was taken from "Leightons Street Map of Auckland City and Suburbs", dated ca.1950 (Leightons 1950). (Reproduced here with permission from Auckland City Libraries.)

As mentioned above, throughout this process of assigning 2006 meshblocks to 1956 census areas, the sizes of the digitally constructed areas were compared with the areas given in the 1956 census publication. It was noticed that the overall area obtained from the meshblocks was around 1% down on the total of the areas given in the census publication. This could easily be due to coastline and inland water being included in some of the 1956 areas but not in the aggregated 2006 meshblocks. Also, it was often not possible to match the sizes of individual census areas more closely without a significantly greater amount of processing of the meshblocks. However, considering the areal interpolation which was to take place later, any errors in the smaller census areas would be unlikely to impact significantly on the final interpolated areas and populations.

Once it was established which of the 2006 meshblocks corresponded to which of the 1956 statistical areas, the meshblocks were merged to give the 1956 areas which could be used in further GIS analyses.

B.2 Tables for Determining Census Populations.

The 1956 New Zealand population census publication (NZDoS 1959) contained maps and tables which were useful in constructing the census geography, as detailed above. It also contained tables of data useful in determining the population distributions. Copies of the following relevant tables are reproduced (at reduced size and resolution) in the following pages of this appendix:

Volume I Table 10 (VI_10): [Split by gender]	page 235
Population of Administrative Counties (Including Dependent Town Districts)	
Volume II Table 9 (VII_9): [Split by gender and age group]	page 237
Counties (Excluding Interior Boroughs and Town Districts)	
Volume II Table 11 (VII_11): [Split by gender and age group]	page 243
(a) Town Districts not forming parts of Counties (Independent Town Districts)	
(b) Town Districts forming parts of Counties (Dependent Town Districts)	

(Note that the copies were made using a book scanner, equivalent to a camera fixed above the open book. This, combined with the thickness of the book, meant that good positioning of the individual pages was difficult to achieve prior to scanning.)

10. POPULATION OF ADMINISTRATIVE COUNTIES

NOTE: Town Districts forming parts of counties have been included with the counties in which they are situated.

COUNTY	Population (Including Maoris)			Population (Excluding Maoris)		
	Males	Females	Total	Males	Females	Total
Mangoni	4,116	3,638	7,754	2,278	1,887	4,165
Whangara	1,233	1,270	2,503	673	644	1,317
Hokitika	3,727	3,598	7,325	1,340	1,260	2,600
Bay of Islands	6,357	5,923	12,280	3,371	2,968	6,339
Whangarei	7,504	6,888	14,392	6,734	5,947	12,681
Hobson	3,599	3,110	6,709	2,928	2,487	5,415
Otago	3,637	3,189	6,826	3,108	2,712	5,820
Rodney	3,013	2,752	5,765	1,524	1,384	2,908
Waitemata	16,764	15,460	32,224	16,302	15,050	31,352
Great Barrier Island	147	124	271	131	104	235
Manukau	11,933	9,746	21,679	11,034	9,192	20,226
Franklin	9,447	7,405	16,852	4,088	3,674	7,762
Raglan	6,337	5,725	12,062	8,441	7,607	16,048
Waikato	8,583	7,676	16,259	4,088	3,674	7,762
Waipa	9,357	8,334	17,691	7,512	6,670	14,182
Otorohanga	4,269	3,632	7,901	2,760	2,465	5,225
Waitemata	5,001	4,178	9,179	3,330	2,821	6,151
Tauaruru	2,303	1,936	4,239	1,427	1,187	2,614
Coromandel	1,432	1,329	2,761	1,263	1,117	2,380
Thames	1,689	1,520	3,209	1,466	1,313	2,779
Hauraki Plains	2,979	2,773	5,752	2,725	2,506	5,231
Ohinemuri	2,317	2,084	4,401	2,017	1,764	3,781
Plato	6,498	5,876	12,374	6,110	5,495	11,605
Matamata	10,619	8,684	19,303	9,460	7,770	17,230
Tauranga	9,801	9,052	18,853	10,432	9,813	20,245
Rotorua	7,707	6,800	14,507	6,861	6,070	12,931
Taupo	2,309	2,115	4,424	4,533	4,306	8,839
Whakatane	7,981	7,115	15,096	4,630	3,928	8,558
Opoitiki	2,672	2,480	5,152	1,813	1,701	3,514
Matakana	969	865	1,834	167	122	289
Uawa	3,184	3,044	6,228	770	636	1,406
Waikohu	899	808	1,707	450	329	779
Cook	2,014	1,721	3,735	1,192	1,000	2,192
Waioa	4,621	4,282	8,903	3,365	3,265	6,630
Hawke's Bay	10,311	9,470	19,781	10,571	9,872	20,443
Waipawa	2,118	1,844	3,962	1,871	1,719	3,590
Patangata	1,910	1,612	3,522	1,530	1,183	2,713
Waipukurau	711	580	1,291	1,329	2,059	3,388
Dannevirke	2,519	2,319	4,838	2,294	2,105	4,399
Woodville	979	861	1,840	937	826	1,763
Ohura	1,010	871	1,881	874	736	1,610
Clifton	1,470	1,302	2,772	1,282	1,136	2,418
Ingleswood	4,683	4,389	9,072	4,395	4,136	8,531
Stratford	1,765	1,556	3,321	1,725	1,536	3,261
Egmont	3,361	3,041	6,402	3,260	2,770	6,030
Eltham	1,923	1,708	3,631	2,785	2,553	5,338
Waimate West	1,556	1,442	2,998	1,873	1,732	3,605
Hawera	3,598	3,245	6,843	2,896	2,820	5,716
Kaitiaki	2,228	1,779	4,007	1,820	1,487	3,307
Waimarino	2,106	1,694	3,800	1,619	1,295	2,914
Waioata	1,452	1,236	2,688	845	700	1,545
Wanganui	1,691	1,691	3,382	1,650	1,537	3,187
Rangitikei	7,811	5,396	13,207	3,614	2,375	5,989
Kiwitea	1,277	1,085	2,362	1,262	1,069	2,331
Pohangina	697	600	1,297	694	596	1,290
Orua	2,411	2,205	4,616	2,099	2,040	4,139
Manawatu	3,848	3,367	7,215	3,146	2,928	6,074
Kairanga	5,627	2,631	8,258	3,388	2,514	5,902
Horowhenua	4,840	4,444	9,284	4,339	3,933	8,272
Hutt	12,120	11,232	23,352	11,842	10,980	22,822
Makara	5,159	5,234	10,393	4,872	4,949	9,821
Pahiatua	1,437	1,340	2,777	1,463	1,303	2,766
Akitio	662	527	1,189	600	484	1,084
Eketahuna	1,040	866	1,906	1,018	844	1,862
Mauriceville	283	263	546	270	254	524
Castlepoint	365	270	635	344	264	608
Masterton	1,828	1,679	3,507	1,611	1,487	3,098
Waipara South	1,673	1,417	3,090	1,573	1,330	2,903
Featherston	2,035	1,752	3,787	1,869	1,629	3,498
TOTALS, NORTH ISLAND COUNTIES ..	284,371	251,996	536,367	232,300	203,933	436,233

10. POPULATION OF ADMINISTRATIVE COUNTIES - continued

NOTE: Town Districts forming parts of counties have been included with the counties in which they are situated

COUNTY	Population (Including Maoris)			Population (Excluding M	
	Males	Females	Total	Males	Females
Sounds	498	453	951	439	405
Marlborough	4,682	4,086	8,768	4,494	3,898
Awatere	919	773	1,692	913	770
Kaikoura	1,673	1,570	3,243	1,635	1,529
Amuri	1,640	1,341	2,981	1,619	1,326
Cheviot	796	670	1,466	791	666
Collingwood	593	490	1,083	579	466
Takaka	1,230	1,005	2,235	1,210	997
Waimea	8,587	8,445	17,032	8,465	8,196
Buller	2,468	2,100	4,568	2,442	2,090
Murchison	821	651	1,472	808	643
Inangahua	767	1,590	2,357	1,574	1,574
Grey	2,754	2,255	5,009	2,717	2,240
Westland	2,437	2,100	4,537	2,368	2,029
Waipara	1,547	1,266	2,813	1,542	1,266
Kowai	1,096	1,004	2,100	1,092	1,001
Ashley	395	297	692	390	296
Rangitara	1,862	1,782	3,644	1,719	1,623
Byre	794	1,626	2,420	828	792
Oxford	801	1,537	2,338	799	835
Tawera	415	339	754	411	338
Malvern	2,856	1,722	4,578	2,842	1,716
Paparua	6,928	6,167	13,095	6,859	6,127
Waimairi	12,465	12,832	25,297	12,424	12,790
Heathcote	2,463	2,765	5,228	2,455	2,746
Halswell	701	553	1,254	701	551
Mount Herbert	309	296	605	281	269
Akaroa	745	657	1,402	718	648
Chatham Islands	283	241	524	186	133
Wairewa	479	436	915	445	410
Springs	1,312	1,086	2,398	1,310	1,082
Ellesmere	1,512	1,363	2,875	1,496	1,351
Selwyn	902	756	1,658	898	755
Ashburton	6,056	5,233	11,289	6,046	5,222
Geraldine	2,838	2,498	5,336	2,777	2,430
Levels	2,775	2,468	5,243	2,757	2,453
Mackenzie	1,664	1,457	3,121	1,661	1,453
Waimate	3,391	3,008	6,399	3,370	2,990
Waitaki	4,373	4,376	8,749	4,333	4,358
Waikamo	573	526	1,099	572	526
Waikouaiti	2,066	1,832	3,898	2,050	1,808
Peninsula	1,873	1,804	3,677	1,851	1,784
Tairā	3,736	3,468	7,204	3,732	3,455
Bruce	2,101	1,857	3,958	2,099	1,851
Clutha	3,394	2,727	6,121	3,357	2,715
Tsapeka	4,325	3,514	7,839	4,282	3,494
Maniototo	1,501	1,232	2,733	1,519	1,228
Vincent	2,588	2,219	4,807	2,565	2,203
Lake	993	819	1,812	992	816
Southland	15,079	12,795	27,874	14,831	12,768
Wallace	5,326	4,492	9,818	5,275	4,459
Fiord	7	4	11	7	4
Stewart Island	314	227	541	251	191
TOTALS, SOUTH ISLAND COUNTIES ..	134,361	119,577	253,938	132,534	118,066
GRAND TOTALS, NEW ZEALAND COUNTIES ..	418,732	371,573	790,305	364,834	321,999

10. POPULATION OF ADMINISTRATIVE COUNTIES

NOTE: Town Districts forming parts of counties have been included with the counties in which they are situated.

COUNTY	Population (Including Maoris)			Population (Excluding Maoris)		
	Males	Females	Total	Males	Females	Total
Mangonui	4,116	3,638	7,754	2,278	4,165	4,165
Whangaroa	1,233	1,270	2,503	673	1,317	1,317
Hokianga	3,787	3,598	7,385	1,340	2,890	2,890
Bay of Islands	6,387	5,923	12,310	3,371	6,239	6,239
Whangarei	7,504	6,888	14,392	6,134	11,681	11,681
Hobson	3,599	3,110	6,709	2,928	5,415	5,415
Otago	3,537	3,189	6,826	3,108	5,820	5,820
Rodney	3,519	3,043	6,532	3,254	6,040	6,040
Waitemata	16,764	15,460	32,224	16,302	31,352	31,352
Great Barrier Island	147	124	271	131	235	235
Mamaku	11,933	9,746	21,679	11,034	20,226	20,226
Franklin	9,447	8,485	17,932	8,441	16,048	16,048
Raglan	6,337	5,725	12,062	4,586	8,674	8,674
Waikato	8,583	7,676	16,259	7,762	14,684	14,684
Waipa	9,357	8,334	17,691	8,371	15,883	15,883
Otorohanga	4,249	3,632	7,881	3,179	5,842	5,842
Waitemata	5,001	4,178	9,179	5,070	9,871	9,871
Tauramuni	2,303	1,936	4,239	1,427	2,614	2,614
Coromandel	1,432	1,329	2,761	1,117	2,380	2,380
Thames	1,669	1,352	3,021	1,466	2,779	2,779
Hauraki Plains	2,979	2,773	5,752	2,725	5,231	5,231
Piako	2,317	2,084	4,401	2,017	3,781	3,781
Chimnani	6,498	5,876	12,374	6,110	11,605	11,605
Matamata	10,619	8,644	19,263	9,480	17,230	17,230
Tauranga	9,801	9,052	18,853	6,861	14,323	14,323
Rotorua	7,707	6,800	14,507	7,280	14,233	14,233
Taupo	5,309	4,307	9,616	3,962	7,744	7,744
Whakatane	7,984	7,115	15,099	4,630	8,558	8,558
Opoitiki	2,672	2,480	5,152	1,041	2,284	2,284
Matukohu	969	865	1,834	829	1,677	1,677
Waipua	3,184	3,044	6,228	770	1,406	1,406
Uawa	899	808	1,707	450	329	329
Waikohu	2,014	1,724	3,738	1,192	2,192	2,192
Cook	4,621	4,282	8,903	3,365	6,473	6,473
Waipa	4,182	3,818	8,000	1,837	3,365	3,365
Hawke's Bay	10,314	9,470	19,784	8,575	16,447	16,447
Waipawa	2,118	1,844	3,962	1,671	3,590	3,590
Patangata	1,910	1,612	3,522	1,530	2,859	2,859
Waipukurua	711	580	1,291	636	1,183	1,183
Dannevirke	2,519	2,319	4,838	2,294	4,399	4,399
Woodville	861	811	1,672	826	1,763	1,763
Ohura	1,010	871	1,881	874	1,610	1,610
Clifton	1,470	1,302	2,772	1,282	2,418	2,418
Taranaki	4,683	4,389	9,072	4,395	8,531	8,531
Inglewood	1,765	1,556	3,321	1,725	3,261	3,261
Stratford	3,361	2,858	6,219	3,260	5,338	5,338
Egmont	3,306	3,041	6,347	2,785	4,930	4,930
Ritanga	2,105	1,708	3,813	1,661	3,307	3,307
Waimea West	1,923	1,708	3,631	1,873	3,534	3,534
Hawera	3,358	3,245	6,603	1,359	2,591	2,591
Patea	2,228	1,779	4,007	2,896	5,716	5,716
Kaitieke	1,452	1,294	2,746	1,487	2,977	2,977
Waimarino	1,805	1,694	3,500	1,619	2,914	2,914
Waitotara	1,891	1,736	3,627	1,070	2,514	2,514
Wanganui	1,907	1,707	3,614	1,537	3,167	3,167
Rangitikei	7,811	5,396	13,207	6,820	11,415	11,415
Kiwitota	1,085	1,065	2,150	1,262	1,069	1,069
Pohangina	697	600	1,297	694	596	596
Oroua	2,411	2,205	4,616	2,099	2,940	2,940
Manawatu	3,848	3,367	7,215	3,626	4,139	4,139
Kairanga	5,627	2,631	8,258	3,627	6,772	6,772
Horowhenua	4,840	4,444	9,284	4,339	5,902	5,902
Ruri	12,120	11,232	23,352	11,840	20,822	20,822
Makara	5,159	5,234	10,393	4,873	9,822	9,822
Pahiatua	1,497	1,340	2,837	1,468	2,771	2,771
Akitio	862	866	1,728	600	1,084	1,084
Eketahuna	1,040	927	1,967	1,018	1,862	1,862
Mauriceville	283	263	546	270	295	295
Castlepoint	1,828	1,679	3,507	1,611	2,544	2,544
Masterton	1,873	1,417	3,290	1,487	3,098	3,098
Wairarapa South	2,035	1,752	3,787	1,869	2,903	2,903
Featherston				1,629	3,498	3,498
TOTALS, NORTH ISLAND COUNTIES	284,371	251,996	536,367	232,300	203,933	436,233

10. POPULATION OF ADMINISTRATIVE COUNTIES - continued

NOTE: Town Districts forming parts of counties have been included with the counties in which they are situated

COUNTY	Population (Including Maoris)			Population (Excluding Maoris)		
	Males	Females	Total	Males	Females	Total
Sounds	498	453	951	439	405	844
Marlborough	4,682	4,086	8,768	4,494	3,898	8,392
Awatere	919	773	1,692	813	770	1,683
Kaikoura	1,673	3,283	4,956	1,529	3,164	4,693
Amuri	1,640	2,981	4,621	1,526	2,945	4,471
Cheviot	670	796	1,466	791	666	1,457
Collingwood	553	1,083	1,636	579	1,045	1,624
Takaka	1,230	1,005	2,235	1,210	977	2,207
Waimea	8,587	8,445	17,032	8,465	16,661	16,661
Buller	2,468	2,100	4,568	2,442	2,090	4,532
Marionston	821	651	1,472	808	643	1,451
Inangahua	1,767	1,590	3,357	1,751	1,574	3,325
Grey	2,754	2,255	5,009	2,717	2,240	4,957
Westland	2,457	2,100	4,557	2,368	2,029	4,397
Waipara	1,547	1,266	2,813	1,542	1,266	2,808
Kowai	1,096	1,004	2,100	1,092	1,001	2,093
Ashley	395	297	692	390	296	686
Rangitara	1,862	1,782	3,644	1,719	1,623	3,342
Lyre	832	794	1,626	828	792	1,620
Oxford	801	836	1,637	799	835	1,634
Tawera	415	359	774	411	348	759
Malvern	2,856	1,722	4,578	2,842	1,716	4,558
Paparoa	6,928	6,167	13,095	6,859	6,127	12,986
Waimairi	12,465	12,832	25,297	12,424	12,790	25,214
Heathcote	2,463	2,765	5,228	2,455	2,746	5,201
Haiawell	853	1,534	2,387	701	1,552	2,253
Mount Herbert	309	605	914	284	550	834
Akaroa	745	657	1,402	718	648	1,366
Cantham Islands	283	241	524	186	133	319
Waipara	479	436	915	445	410	855
Springs	1,312	1,086	2,398	1,310	1,082	2,392
Ellesmere	1,512	1,363	2,875	1,496	1,311	2,807
Selwyn	902	756	1,658	898	755	1,653
Ashburton	6,096	5,233	11,289	6,046	5,222	11,268
Geraldine	2,498	2,498	4,996	2,477	2,430	4,907
Levels	2,775	2,468	5,243	2,757	2,453	5,210
Mackenzie	1,664	1,457	3,121	1,661	1,453	3,114
Waimate	3,391	3,008	6,399	3,370	2,990	6,360
Waikato	4,973	4,376	9,349	4,933	4,358	9,291
Waikouaiti	573	526	1,099	572	526	1,098
Peninsula	2,066	1,832	3,898	2,050	1,808	3,858
Tairā	1,873	1,804	3,677	1,851	1,784	3,635
Bruce	3,736	3,468	7,204	3,732	3,455	7,187
Clutha	1,857	1,857	3,998	1,851	1,851	3,950
Tuapeka	2,101	2,727	4,828	2,715	3,502	6,217
Maniototo	2,727	2,727	5,454	2,717	2,717	5,434
Vincent	1,232	1,232	2,755	1,232	1,232	2,747
Lake	2,588	2,219	4,807	2,565	2,203	4,768
Southland	993	819	1,812	992	816	1,808
Wallace	15,079	12,795	27,874	14,831	12,768	27,599
Piort	5,326	4,492	9,818	5,275	4,459	9,734
Stewart Island	7	4	11	7	4	11
TOTALS, SOUTH ISLAND COUNTIES	134,361	119,577	253,938	132,534	118,066	250,600
GRAND TOTALS,						
NSW ZEALAND COUNTIES	418,732	371,573	790,305	364,834	321,999	686,833

9. AGE: DISTRIBUTION BY COUNTIES - MALES - continued

Table with columns for County (Exclusive of Interior Boroughs and Town Districts) and Age in Years (Under 5 to 55 and over). Rows list various counties like Eyre, Oxford, Tavers, etc., and include a totals row for South Island counties and New Zealand.

Volume II Table 9 (VII_9)

9. AGE: DISTRIBUTION BY COUNTIES - MALES - continued

Table with columns for County (Exclusive of Interior Boroughs and Town Districts) and Age in Years (55 and under 60 to 90 and over, Unspecified Adults, and Minors). Rows list various counties like Eyre, Oxford, Tavers, etc., and include a totals row for South Island counties and New Zealand.

Volume II Table 9 (VII_9)

9. AGE: DISTRIBUTION BY COUNTIES - FEMALES

Table with 17 columns for age groups (Under 5 to 50 and under 55) and rows for various counties including Mangonui, Whangaroa, Hokianga, Bay of Islands, Whangarei, Hobson, Otamatea, Rodney, Waitemata, Great Barrier Island, Manukau, Franklin, Napier, Waikato, Waipa, Otorohanga, Waikato, Tamarunui, Coromandel, Thames, Hauraki Plains, Ohinemuri, Piako, Matamata, Tauranga, Rotorua, Taupo, Whakatane, Opotiki, Matakaoa, Waipoua, Uawa, Waikohu, Cook, Wairoa, Hawke's Bay, Waipawa, Patangata, Waipukurau, Dannevirke, Woodville, Chura, Clifton, and Taranaki.

Volume II Table 9 (VII_9)

9. AGE: DISTRIBUTION BY COUNTIES - FEMALES - continued

Table with 17 columns for age groups (55 and under 60 to 90 and Over), 2 columns for Unspecified (Adults, Minors), and 3 columns for Total (Total Adults, Total Minors, Grand Total). Rows are identical to the previous table, listing counties from Mangonui to Taranaki.

Volume II Table 9 (VII_9)

11. AGE: DISTRIBUTION BY TOWN DISTRICTS - MALES

Town District	Age, in Years													
	Under 5	5 and under 10	10 and under 14	14	15	15 and under 20	20	21 and under 25	25 and under 30	30 and under 35	35 and under 40	40 and under 45	45 and under 50	50 and under 55
	(a) Town Districts not forming parts of Counties													
Kawakawa	52	45	29	5	8	17	5	12	23	25	27	22	28	16
Hikurangi	70	67	55	10	8	33	7	27	37	21	27	30	29	27
Kamo	58	53	31	13	10	19	3	30	42	40	21	44	32	37
Onerahi	112	81	40	12	14	21	10	32	74	71	42	38	42	29
Warkworth	48	57	31	7	6	22	2	26	32	42	35	30	25	27
Leamington	70	58	56	8	15	38	5	28	49	35	33	45	27	24
Murupara	92	52	21	5	5	22	9	55	89	55	40	30	21	19
Ohura	48	39	29	4	3	8	2	22	19	23	22	25	28	13
Mennis	44	32	33	4	3	17	2	14	22	19	25	32	23	19
Waverley	60	38	29	5	12	16	5	14	35	30	28	24	28	19
Mamunui	71	51	30	11	10	22	5	32	38	32	30	26	28	22
Mangaweka	25	23	16	1	2	3	2	12	13	18	8	9	6	11
Hunterville	26	40	18	6	5	8	3	9	22	19	20	27	19	12
Bulls	46	39	32	12	9	23	4	16	37	30	14	28	25	26
Totals, males, North Island independent town districts ..	822	695	450	103	110	269	64	329	532	460	372	410	361	301
Takaka	49	48	29	10	3	11	2	12	24	30	32	26	22	22
Leaton	53	45	26	5	6	23	3	10	24	23	23	25	28	18
Wyndham	47	32	20	5	5	12	4	18	19	27	21	20	25	12
Lusden	57	43	9	6	1	11	5	23	22	24	23	22	13	15
Nightcaps	42	52	21	6	5	14	2	18	26	24	20	24	26	14
Otutau	37	48	27	12	6	23	2	14	26	31	22	31	25	11
Totals, males, South Island independent town districts ..	285	268	132	44	26	94	18	95	141	159	141	148	139	92
Totals, males, all independent town districts ..	1,107	963	582	147	136	363	82	424	673	619	513	558	500	393
(b) Town Districts forming parts of Counties (Parent Counties shown in parentheses)														
Kohukohu (Hokianga)	13	19	7	2	2	3	3	7	6	7	7	10	5	5
Rawene (Hokianga)	29	28	23	5	5	7	1	6	15	17	13	10	14	11
Russell (Bay of Islands)	24	27	20	1	5	10	3	11	20	17	21	23	23	17
Mercer (Franklin)	24	16	9	3	3	6	3	16	17	15	18	12	15	4
Te Kawhata (Waikato)	49	51	18	5	5	16	3	15	23	24	22	30	27	20
Ohupo (Waipa)	24	19	5	4	5	10	2	10	13	13	7	8	17	5
Kihikihiki (Waipa)	47	61	37	9	5	8	1	11	32	36	18	24	23	21
Te Karaka (Waikohu)	36	25	13	5	5	14	4	10	21	14	7	7	10	10
Patutahi (Cook)	16	15	8	1	2	6	1	5	8	9	8	9	4	8
Kaponga (Eltham)	39	33	24	1	3	19	3	12	23	19	19	20	8	11
Normanby (Hawera)	28	36	19	6	6	15	2	11	14	13	11	18	14	9
Totals, males, North Island dependent town districts ..	329	330	183	42	46	114	36	114	192	182	151	171	160	121
Havelock (Marlborough)	18	16	7	3	4	6	..	2	10	13	12	4	9	2
Southbridge (Ellesmere)	28	30	10	3	7	11	1	8	17	14	18	12	12	6
Outram (Taieri)	17	17	13	5	4	6	1	6	13	11	9	14	10	8
Edendale (Southland)	44	24	15	3	5	10	3	8	14	23	16	21	25	9
Totals, males, South Island dependent town districts ..	107	87	45	14	20	33	5	24	54	61	55	51	56	25
Totals, males, all dependent town districts ..	436	417	228	56	66	147	31	138	246	243	206	222	216	146
Totals, males, all town districts	1,543	1,380	810	203	202	510	113	562	919	862	719	780	716	539

11. AGE: DISTRIBUTION BY TOWN DISTRICTS - MALES - continued

Town District	Age, in Years										Total Adults	Total Minors	Grand Total
	55 and under 60	60 and under 65	65 and under 70	70 and under 75	75 and under 80	80 and under 85	85 and under 90	90 and Over	Unspecified Adults	Unspecified Minors			
	(a) Town Districts not forming parts of Counties												
Kawakawa	12	8	10	7	3	..	1	197	161	358
Hikurangi	36	16	14	15	11	3	3	278	250	528
Kamo	14	13	16	13	10	3	2	317	167	594
Onerahi	32	17	31	15	9	3	4	1	440	290	730
Warkworth	15	17	10	8	11	2	2	1	282	173	455
Leamington	25	22	24	20	9	6	1	348	259	598
Murupara	9	2	1	2	323	206	529
Ohura	13	6	3	7	2	1	1	184	134	318
Mauia	12	12	17	13	5	4	217	135	352
Waverley	16	15	11	13	2	5	2	242	185	427
Mamunui	20	13	12	9	11	3	1	..	1	..	276	200	478
Mangaweka	2	6	6	3	2	2	96	72	170
Hunterville	12	6	10	8	3	2	1	2	172	106	278
Bulls	17	12	8	5	15	2	235	165	400
Totals, males, North Island independent town districts ..	217	165	172	136	93	40	18	4	1	1	3,611	2,514	6,125
Takaka	17	14	8	10	4	1	1	..	223	152	375
Leaton	10	8	21	17	5	2	214	161	375
Wyndham	10	8	12	10	7	3	2	194	125	319
Lusden	6	8	9	7	6	2	..	2	184	132	316
Nightcaps	11	5	10	5	4	187	142	329
Otutau	15	14	10	14	6	1	2	222	155	377
Totals, males, South Island independent town districts ..	69	57	70	63	32	9	6	..	3	..	1,224	867	2,091
Totals, males, all independent town districts ..	286	222	242	199	125	49	24	4	4	1	4,835	3,381	8,216
(b) Town Districts forming parts of Counties (Parent Counties shown in parentheses)													
Kohukohu (Hokianga)	3	3	3	7	4	..	1	68	49	117
Rawene (Hokianga)	4	7	5	7	4	1	1	113	98	211
Russell (Bay of Islands)	29	13	18	13	5	7	1	238	90	308
Mercer (Franklin)	13	4	1	2	..	1	1	119	64	183
Te Kawhata (Waikato)	10	5	7	5	3	1	192	147	339
Ohupo (Waipa)	5	6	4	5	..	3	1	97	69	166
Kihikihiki (Waipa)	11	9	8	10	7	..	3	213	168	381
Te Karaka (Waikohu)	5	6	7	4	6	1	2	110	102	212
Patutahi (Cook)	3	6	6	6	4	1	2	79	49	128
Kaponga (Eltham)	2	8	2	6	4	134	122	256
Normanby (Hawera)	8	6	13	6	6	1	..	1	1	..	132	112	244
Totals, males, North Island dependent town districts ..	93	73	74	71	43	16	12	1	1	..	1,475	1,070	2,545
Havelock (Marlborough)	3	8	10	10	4	87	54	141
Southbridge (Ellesmere)	11	11	11	11	2	4	1	138	90	228
Outram (Taieri)	14	10	13	6	5	4	123	63	186
Edendale (Southland)	14	9	12	12	4	2	1	170	104	274
Totals, males, South Island dependent town districts ..	42	38	46	39	15	10	2	518	311	829
Totals, males, all dependent town districts ..	135	111	120	110	58	26	14	1	1	..	1,993	1,381	3,374
Totals, males, all town districts	421	333	362	309	183	75	38	5	5	1	6,828	4,762	11,590

Volume II Table 11 (VII_11)

Volume II Table 11 (VII_11)

Appendix C ArcGIS Python Scripts

An outline of how the ArcGIS ModelBuilder and Python scripts were used is given in section C.1. Section C.2 lists the main scripts used in this study.

C.1 Processing with ArcGIS ModelBuilder and Python

Python scripts were used to automate, or semi-automate, most of the GIS processes carried out in this study. Steps discussed here refer to those illustrated in Figure 4.6.

The initial scripts for creating the latitude grids and then processing the census areas were combined within the ArcGIS ModelBuilder. The ModelBuilder provides an environment for combining python scripts and ArcGIS Tools in an easy to use graphical interface. Input data and parameters can be specified as illustrated in Figure C.1 which shows a simple model equivalent to the three processes in the top right of the workflow in Figure 4.6. However, if greater control is required over the checking for and the naming of input and output parameters, and intermediate data tables, then the model obtained is closer to that shown in Figure C.2 for the same three processes. Given that these three processes were only a small section of the overall workflow, as depicted in Figure 4.6, it was decided to limit the use of the ModelBuilder and to concentrate instead on programming most of the control logic within the python scripts. In this way it was also much easier to use a naming convention for the various tables and data sets which reflected the input parameters used.

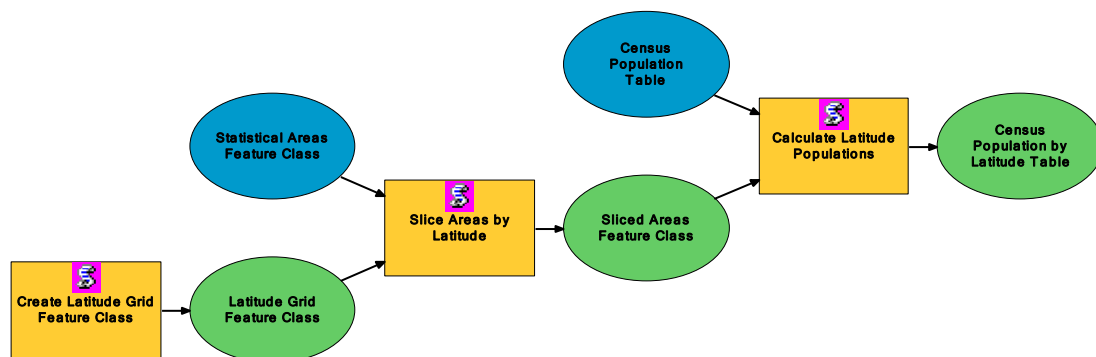


Figure C.1 A simple ArcGIS ModelBuilder model for processing census data

Python scripts are shown in yellow with the 'script' symbol. Input data is shown in blue and output data for the different scripts is shown in green. Arrows indicate the direction of process and data flow. Each of the scripts shown could, alternatively, be an ArcGIS Tool or another model comprising any number of linked processes.

To illustrate the naming convention which was used, consider output data from the final age standardisation process (described in section 4.3.2), which depends on: the year of the census data used; the standard population; the latitude slice size; the residence criterion; the life period of interest; gender; and any other pre-filter or factor being considered. Gender information was included within the table by storing data for male, female and 'all', and the two parent directories of the ArcGIS geodatabase were named after any pre-filter and the factor type. All other information was then included in the name of the table as illustrated in Table C.1.

Despite the limitation of the ModelBuilder mentioned above, it does have the advantage of being able to generate a python script for the processes in the model. This basic script could then be used as a starting point with more advanced control logic added to the script later, by hand. Although the model depicted in Figure C.1 links three of the main processes in the workflow of Figure 4.6, most of the steps indicated in Figure 4.6 were run separately. ArcGIS provides a 'batch' ability so that scripts can be run repeatedly with various input parameters. Running each process in turn in batch mode for a number of different sets of input parameters was found to be more efficient than following a given set of parameters through the full workflow, and then repeating the workflow for each set of parameters.

The initial processes, in column (a) of Figure 4.6, involved reformatting the various input data into either Microsoft Excel tables or ArcGIS geodatabase tables manually. However, apart from generating new tables for any pre-filters or factors, these processes only needed to be carried out once. The next set of processes, in column (b) of Figure 4.6, were predominantly involved in selecting the relevant data from the other tables and carrying out a certain amount of pre-processing. A latitude grid with a single slice covering the extent of New Zealand was also used to determine the residency of MS cases. This data flow in the workflow process is indicated with a dashed line. The top two processes in column (c) of Figure 4.6 carried out the areal interpolation depicted in Figure 4.5 and, again, once these had been carried out for the required choices of latitude slice size and year, they did not need to be repeated. The main processes which did need to be repeated are those dependent on different residence and life periods, pre-filters and factors. However, assuming that the prerequisite processes had already been carried out, the inclusion of a new pre-filter

Convention for normalised (standardised) data:		
directory:	pre_factor\ factor_type\ ND.mdb nyyyypppp_ss_rrrlll_f	
directory:		
geodatabase:		
table:		
Example:		
directory:	res_all\ ms_type\ ND.mdb n1956euro_10_all0a4_s	
directory:		
geodatabase:		
table:		
Code	e.g.	Description
YYYY	1956	Year used for census data (1956 or 2006)
pppp	euro	Standard population (segi or euro)
ss	10	Latitude slice size (units of 0.1°) 01: 0.1°; 05: 0.5°; 10: 1.0°; 20: 2.0°; etc.
rrr	all	Period of residence within New Zealand all: from birth to onset; prv: only at prevalence 2006; 0a4: age 0–4; 4o0: 0–4 years prior to onset; etc.
lll	0a4	Life period (must be a sub-period of residence period) all: from birth to onset; prv: prevalence 2006; 0a4: age 0–4; 4o0: 0–4 years prior to onset; etc.
f	s	Factor (depends upon factor type), e.g. for ms_type a: all; p: PPMS; r: RPMS; s: SPMS; x: unspecified

Table C.1 The naming convention used for the output data

The naming convention used for the normalised (standardised) output data is indicative of the conventions used for all of the data output.

or factor would simply require three further processes to be run, for each of the new sets of parameters: the selection of the relevant data, the age standardisation process, and the data analysis.

The factor tables, required for the selection process, were generated manually within ArcGIS. Each MS case had an entry in the table with a column for the factor which was a designated single character. For example, for the factor ms_type, corresponding to MS phenotype, the factor could take any of the codes given in the last line of Table C.1, except for 'a' which was always reserved for 'all' in output tables. The pre-filter tables were similar to the factor tables except that only cases corresponding to a particular factor would be listed. If a pre-filter was applied to the selection then only cases corresponding to that factor would be selected. The selected data would then be stratified according to the factor before being stored for further processing.

The main process in the workflow, that of age-standardising the data, takes the selected, stratified MS data and combines it with the processed census data to give data by latitude in addition to the stratification by factor. Details of the age standardisation process are discussed in section, 4.3.2. Note that the output data stored from this process is the main output data of this study.

In order to summarise the output data generated from the age standardisation process, an analysis process is also available to systematically produce graphs and fit data in a pdf report format. Details of the analyses are given in sections 4.3.3 and 4.3.4. These graphs and output statistics can be scanned quickly to establish salient features of the output data. Data which requires a closer investigation or further analysis can be extracted from the ArcGIS tables in formats suitable either for Microsoft Excel or for R. Similarly, data required for final reporting procedures can be extracted and processed manually as required.

[The next section continues on the next page.]

C.2 ArcGIS Python Scripts

The following ArcGIS python scripts are listed on the following pages:

Script	Process	Page
CreateLatitudeGrid.py	Create Latitude Grid at Required Resolution	251
SelectAreas.py	Select Census Statistical Areas by Year	253
SelectPopData.py	Select Census Population Data by Age Group and Gender	255
SelectEpochData.py	Select MS Data by NZ Residence, Age or Life Period, Gender, Pre-Filter and Factor	259
SliceAreas.py	Slice Census Statistical Areas by Latitude Grid	273
LatitudePops.py	Re-aggregate Population Data by Latitude Sliced Statistical Areas	275
ProcessEpochLats.py	Standardise MS Data by Latitude, Gender, and Factor	279
LatitudeFunctions.py	Utility Functions	285
LatitudeDataNames.py	Utility Functions	288
BatchFactors.py	Batch Processing	296
PlotGraphsX4.py	Systematic Data Analysis and Statistics	298
PlotFunctions.py	Plot Utility Functions	310
TabsToXL.py	Extract Tables to Sheets in Excel File	316
ExtractData2CSV.py	Extract and Combine Tables to one CSV File	318

Table C.2 List of python scripts

The python scripts listed in this section, with page numbers.

Note:

Python uses the indentation of lines to signify a block of code and statements continuing onto more than one line either have a '\ ' preceding the line break or the line break should fall within any parentheses or brackets which are used within the statement. In order to format the scripts for inclusion here, text wrapping has been avoided and long statements have been split according to Python syntax.

CreateLatitudeGrid.py

```
# -----
# CreateLatitudeGrid.py
# -----

# Import system modules
import sys, string, os, arcgisscripting

import LatitudeFunctions as LF
import LatitudeDataNames as LDN

# Create the Geoprocessor object
gp = arcgisscripting.create(9.3)

def main(*argv):
    strCodeSec = "getting parameters"
    strExtraInfo = ""
    try:

        # Set the Geoprocessing environment...
        gp.scratchWorkspace = "D:/WorkSpace"

        # Script arguments...
        iParam=0
        in_GridWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_GridWDS == '#':    in_GridWDS = ""
        iParam=iParam+1
        in_Slice = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_Slice == '#':    in_Slice = "0"
        iParam=iParam+1
        in_Top = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_Top == '#':    in_Top = "-32"
        iParam=iParam+1
        in_Bottom = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_Bottom == '#':    in_Bottom = "-48"
        iParam=iParam+1
        in_Left = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_Left == '#':    in_Left = "164"
        iParam=iParam+1
        in_Right = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_Right == '#':    in_Right = "180"
        iParam=iParam+1
        in_SegLen = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_SegLen == '#':    in_SegLen = "0.1"
        iParam=iParam+1
        in_ScratchWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_ScratchWDS == '#':    in_ScratchWDS = ""
        iParam=iParam+1
        in_KTO = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_KTO == '#':    in_KTO = ""
        iParam=iParam+1

        blnKTO = (in_KTO.lower() == "true")

        strCodeSec = "local variables"

        strExtraInfo = "; floating: ' + '", '".join(
            in_Slice,in_SegLen,in_Top,in_Bottom,in_Left,in_Right) + "'
        dSlice, dSeg, dTop, dBottom, dLeft, dRight = map(float,
            (in_Slice, in_SegLen, in_Top, in_Bottom, in_Left, in_Right))

        strSlice = LF.parseLat(in_Slice)

        # local variables...
        strLatN = LDN.FieldLatN() # "Lat_N"
        strLatS = LDN.FieldLatS() # "Lat_S"
        strLatC = LDN.FieldLatC() # "Lat_C"

        strGridName = LDN.KeepLatGrid(strSlice)
        strGrid = in_GridWDS + "/" + strGridName

        ## Create an empty Feature Class...
        strExtraInfo = "; creating feature class: %s ; %s" % (in_GridWDS, strGridName)
        gp.CreateFeatureclass_management(in_GridWDS, strGridName, "POLYGON", "")

        ## Process: Add Field...
        strExtraInfo = "; adding field %s to %s" % (strLatN, strGrid)
        gp.AddField_management(strGrid, strLatN, "double")

        ## Process: Add Field...
        strExtraInfo = "; adding field %s to %s" % (strLatS, strGrid)
        gp.AddField_management(strGrid, strLatS, "double")

        ## Process: Add Field...
        strExtraInfo = "; adding field %s to %s" % (strLatC, strGrid)
        gp.AddField_management(strGrid, strLatC, "double")

        ## Add the Latitude Grid, building it manually...
```

```

strCodeSec = "build grid"

# Calculate number of points on polys and number of slices
strExtraInfo = "; first calcs..."
iPrec = 3
if strSlice == "00":
    dSlice = (dTop - dBottom)
    nSlc = 1
else:
    nSlc = int(round(((dTop - dBottom) / dSlice), 0))
dHSlice = dSlice * 0.5
nPts = int(round(((dRight - dLeft) / dSeg), 0)) + 1

# Calculate point coordinates
strExtraInfo = "; calc coords..."
vPts1 = [(round(i * dSeg + dLeft, iPrec), 0) for i in range(0, nPts)]
vPts2 = [(round(i * dSeg + dLeft, iPrec), dSlice) for i in range(nPts - 1, -1, -1)]
vPts0 = []
vPts0.extend(vPts1)
vPts0.extend(vPts2)
vPts = [(vPts0[i][0], vPts0[i][1] + dBottom + j * dSlice) for i in range(0, 2*nPts)]
        for j in range(0, nSlc)]
vLat = [(dBottom + j * dSlice, dBottom + j * dSlice + dHSlice, dBottom + (j+1)*dSlice)
        for j in range(0, nSlc)]

vFld = [strLatS, strLatC, strLatN]

# Open an insert cursor for the feature class
strExtraInfo = "; getting cursor & objects..."
pCur = gp.InsertCursor(strGrid)

# Create array and point objects
aLine = gp.CreateObject("Array")
pPt = gp.CreateObject("Point")

strCodeSec = "main loop"
for j in range(0, nSlc):

    strExtraInfo = "; new row j = %d" % j
    # Create a new row, or feature, in the feature class
    pFeat = pCur.NewRow()

    strExtraInfo = "; add points j = %d" % j
    # Add the points to the array
    aLine.RemoveAll()
    for i in range(0, 2*nPts):
        pPt.X, pPt.Y = vPts[j][i][0], vPts[j][i][1]
        aLine.Add(pPt)

    strExtraInfo = "; set geom j = %d" % j
    # Set the geometry of the new feature to the array of points
    pFeat.Shape = aLine

    strExtraInfo = "; set vals j = %d" % j
    # Set the field values
    for i in range(0, 3):
        pFeat.SetValue(vFld[i], vLat[j][i])

    strExtraInfo = "; insert row j = %d" % j
    # Insert the feature
    pCur.InsertRow(pFeat)

print gp.GetMessages()

### Done ###
gp.AddMessage("Done.")
strCodeSec = "End!"
strExtraInfo = ""

if blnKTO:
    strCodeSec = "exception to keep tool open!"
    exit() # keep tool open in case works and gets to here!

except:
    gp.AddMessage("Got exception in CreateLatitudeGrid, section = %s %s"
                  % (strCodeSec, strExtraInfo))
    raise
else:
    return 0 # exit errorlessly

if __name__ == '__main__':
    main(*sys.argv[1:])

```


SelectAreas.py

```
# -----
# SelectAreas.py
# -----

# Import system modules
import sys, string, os, arcgisscripting

import LatitudeFunctions as LF
import LatitudeDataNames as LDN

# Create the Geoprocessor object
gp = arcgisscripting.create(9.3)

def main(*argv):
    strCodeSec = "getting parameters"
    strExtraInfo = ""
    try:

        # Script arguments...
        iParam=0
        in_StatAreas = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_StatAreas == '#':    in_StatAreas = ""
        iParam=iParam+1
        in_AreaField = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_AreaField == '#':    in_AreaField = ""
        iParam=iParam+1
        in_URField = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_URField == '#':    in_URField = ""
        iParam=iParam+1
        in_UrbanCrit = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_UrbanCrit == '#':    in_UrbanCrit = ""
        iParam=iParam+1
        in_CYear = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_CYear == '#':    in_CYear = ""
        iParam=iParam+1
        in_AreasWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_AreasWDS == '#':    in_AreasWDS = LDN.WDSAreas()
        iParam=iParam+1
        in_ScratchWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_ScratchWDS == '#':    in_ScratchWDS = LDN.WDSScratch()
        iParam=iParam+1
        in_KTO = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_KTO == '#':    in_KTO = ""
        iParam=iParam+1

        blnKTO = (in_KTO.lower() == "true")

        if not gp.Exists(in_StatAreas):
            gp.AddMessage("Input Statistical Areas feature class %s does not exist!") % in_StatAreas
            return 0

        strCodeSec = "local variables"
        # local variables...
        strCYear = LF.parseCYear(in_CYear)

        # Feature Class or Table names
        strAreasName = LDN.KeepStatAreas(strCYear)
        strUrbanName = LDN.KeepStatUrban(strCYear)
        strTmpAreasName = "tmpAreas"
        strTmpUrbanName = "tmpUrban"

        # ... and qualified with Feature Dataset or Workspace
        strAreas = in_AreasWDS + "/" + strAreasName
        strUrban = in_AreasWDS + "/" + strUrbanName
        strTmpAreas = in_ScratchWDS + "/" + strTmpAreasName
        strTmpUrban = in_ScratchWDS + "/" + strTmpUrbanName

        strTmpLayer = "tmpLayer"

        strFld_sa = LDN.FieldStatArea()
        strFld_sqkm = LDN.FieldSqKm()
        strFld_urb = LDN.FieldUrban()

        # Set the scratch workspace
        gp.Workspace = in_ScratchWDS

        strCodeSec = "copying areas"
        gp.CopyFeatures_management(in_StatAreas, strTmpAreas)
        strCodeSec = "adding fields"
        strExtraInfo = "; %s" % strFld_sqkm
        gp.AddField_management(strTmpAreas, strFld_sqkm, "DOUBLE")
        strExtraInfo = "; %s" % strFld_urb
        gp.AddField_management(strTmpAreas, strFld_urb, "TEXT", 0, 0, 5)

        strCodeSec = "calculating fields"
        strExtraInfo = "; %s" % strFld_sqkm
        strVB = "Dim Output as double\\nDim pArea as Iarea\\nSet pArea = [shape]\\n" \
```

```

        + "Output = pArea.area/1000000.0\\n"
gp.CalculateField_management(strTmpAreas, strFld_sqkm, "Output", "VB", strVB)

strExtraInfo = "; %s" % strFld_urb
strUrbanCrit = in_UrbanCrit.replace("'",'\')
strVB = "Dim Output as string\\nif ([\" + in_URField + \"] \" + strUrbanCrit \
+ \") then\\nOutput = \"urban\"\\nelse\\nOutput = \"rural\"\\nend if\\n"
gp.CalculateField_management(strTmpAreas, strFld_urb, "Output", "VB", strVB)

# use layer to tidy up fields and names
fields = gp.ListFields(strTmpAreas)
strFieldInfo = "%s %s VISIBLE NONE" % (in_AreaField, strFld_sa)

strCodeSec = "setting field info"
for fld in fields:
    strExtraInfo = "; %s" % fld
    if (fld.type.upper not in ("GEOMETRY", "OBJECTID")) \
        and (fld.name not in (in_AreaField, strFld_sa)):
        sVis = "VISIBLE" if fld.name in (strFld_urb, strFld_sqkm, ) else "HIDDEN"
        strFieldInfo = "%s;%s %s %s NONE" % (strFieldInfo, fld.name, fld.name, sVis)

strCodeSec = "making layer"
strExtraInfo = ""
gp.MakeFeatureLayer_management(strTmpAreas, strTmpLayer, "", in_ScratchWDS, strFieldInfo)

# now we're done, create copy to keep
strCodeSec = "saving layer copy"
gp.CopyFeatures_management(strTmpLayer, strAreas)

strCodeSec = "getting urban areas"
strExtraInfo = "; selecting"
# select urban areas
strSelTypeByAttrib = "NEW_SELECTION"
whereUrban = "[%s] = 'urban'" % strFld_urb
gp.SelectLayerByAttribute_management(strTmpLayer, strSelTypeByAttrib, whereUrban)

strExtraInfo = "; dissolving"
# dissolve to give single multipart feature
gp.Dissolve_management(strTmpLayer, strTmpUrban,
    strFld_urb, "", "MULTI_PART", "DISSOLVE_LINES")

strExtraInfo = "; saving"
# create copy to keep
gp.CopyFeatures_management(strTmpUrban, strUrban)

### Done ###
gp.AddMessage("Done.")
strCodeSec = "End!"
strExtraInfo = ""

if blnKTO:
    strCodeSec = "exception to keep tool open!"
    exit() # keep tool open in case works and gets to here!

except:
    gp.AddMessage("Got exception in SelectAreas, section = %s %s" % (strCodeSec, strExtraInfo))
    raise
else:
    return 0 # exit errorlessly

if __name__ == '__main__':
    main(*sys.argv[1:])

```

SelectPopData.py

```
# -----
# SelectPopData.py
# -----

# Import system modules
import sys, string, os, win32com.client, arcgisscripting
# still uses win32com for excel instead of xl utils etc

import LatitudeFunctions as LF
import LatitudeDataNames as LDN

# Create the Geoprocessor object
gp = arcgisscripting.create(9.3)

def main(*argv):
    strCodeSec = "getting parameters"
    strExtraInfo = ""
    try:

#     gp.AddMessage("SelectPopData got args:\n%s" % repr(argv))

        # read data from XL 'fields' sheet to find out what data is where
        # - need to know 'Area' field(s) and 'Data' field(s)
        # select according to input parameters: gender, age group, year etc
        # given list of sheets and fields, read in relevent data
        # combine input fields to produce single output field suitably scaled
        # construct table of output data and store in gdb

        # Script arguments...
        iParam=0
        in_XLFile = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_XLFile == '#':     in_XLFile = ""
        iParam=iParam+1
        in_Year = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_Year == '#':     in_Year = ""
        iParam=iParam+1
        in_PopWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_PopWDS == '#':   in_PopWDS = LDN.WDSPop()
        iParam=iParam+1
        in_ScratchWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_ScratchWDS == '#': in_ScratchWDS = LDN.WDSScratch()
        iParam=iParam+1
        in_KTO = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_KTO == '#':     in_KTO = ""
        iParam=iParam+1

        blnKTO = (in_KTO.lower() == "true")

        if not gp.Exists(in_XLFile):
            gp.AddMessage("Input Excel file %s does not exist!") % in_XLFile
            return 0

        strCodeSec = "local variables"
        strExtraInfo = ""
        # Local variables...

        in_AgeGroup = "all"

        strCYear = LF.parseCYear(in_Year)
        allGs = ['a', 'f', 'm']
        (strPer, strPerd, (reqMinAge, reqMaxAge), reqMonths) = LF.parsePeriod(in_AgeGroup)

        # census year groups; last group open ended
        # group definitions, year limits, month limits
        allCYLims = LDN.AllCYearLimits(strCYear)
        numCYs = [ allCYLims[i][1] - allCYLims[i][0] + 1 for i in range(0,len(allCYLims) - 1) ]
        numCYs.append(None)
        lutCYGrp = []
        MaxCYGrp = len(allCYLims) - 1
        # lut for years to groups - not too many years
        for i in range(0,len(allCYLims) - 1):
            lutCYGrp.extend([i] * (numCYs[i]))
        lutCYGrp.extend([len(allCYLims) - 1])
        MaxCYlut = len(lutCYGrp) - 1 # should = allCYLims[len(allCYLims)-1][0]

        # Feature Class or Table names
        strPopTabName = LDN.KeepPop(strCYear)
        strTmpPopTabName = "tmpPop"

        # ... and qualified with Feature Dataset or Workspace
        strPopTab = in_PopWDS + "/" + strPopTabName
        strTmpPopTab = in_ScratchWDS + "/" + strTmpPopTabName

        # other model variables...

        strAreaField = LDN.FieldSPopArea() # "Stat_Area"
        # field name for each gender and all census year groups
```

```

strPopField = {}
for g in allGs: strPopField[g] = LDN.AllFieldsSamplePop(g, strCYear)
strFieldsSheet = "fields"
rangeFlds = ""
data = {}

gp.AddMessage("Examining Excel file...")
strCodeSec = "setting up basic where clause"

# Set the scratch workspace
gp.Workspace = in_ScratchWDS

strExtraInfo = "; year"
# year is straightforward
whereYear = " [Year]=%s " % (strCYear)

# parsed age group for min and max ages, or total
strExtraInfo = "; parsing age groups"

# if in_AgeGroup.strip().lower()=="all" or in_AgeGroup.strip()=="":
if strPerd=="all":
    whereAge = " [Type]='AGE' "
    strAgeGroup = "all"
    cyGrpLo = 0
    cyGrpHi = MaxCYGrp
else:
# to find ranges which overlap compare cross terms i.e compare min with max
# need to filter for high end open
    reqMinAge = int(reqMinAge)
    whereAge = " [Type]='AGE' AND ( [MaxAge] >= %d " % reqMinAge
    cyGrpLo = lutCYGrp[reqMinAge]

    if reqMaxAge=="":
        strAgeGroup = "open"
# but we need an upper age for the ageFac...
# assume upper limit is max of 105 and (MinAge+5) for field
# will need to change if stds get older!
        maxOpenAge = 105
        whereAge = " %s OR [MinAge] >= %d ) " % (whereAge, reqMinAge)
        cyGrpHi = MaxCYGrp
    else:
        reqMaxAge = int(reqMaxAge)
        strAgeGroup = "normal"
        whereAge = " %s AND [MinAge] <= %d ) " % (whereAge, reqMaxAge)
        cyGrpHi = lutCYGrp[reqMaxAge]

strCodeSec = "looping over genders"
for strGender in allGs:
    strExtraInfo = "; gender %s" % strGender
    # fields gender is MALE, FEMALE or blank
    if strGender=="a": whereGen = " ([Gender] IS NULL OR [Gender]='') "
    else: whereGen = " [Gender]='%s' " % ("MALE" if strGender=="m" else "FEMALE")

    whereBase = " WHERE %s AND %s" % (whereYear, whereGen)
    whereAll = "%s AND %s" % (whereBase, whereAge)

    data[strGender] = {}

    strExtraInfo = "; %s; win32com" % strGender
    DSN= ""
        PROVIDER=Microsoft.Jet.OLEDB.4.0;
        DATA SOURCE=%s;
        Extended Properties='Excel 8.0;READONLY=true;IMEX=1';
        "" % in_XLFile
    conn = win32com.client.Dispatch("ADODB.Connection")
    conn.Open(DSN)
    rs = win32com.client.Dispatch("ADODB.Recordset")
    strXLFile = "Excel 8.0;HDR=YES;IMEX=1;Database=%s;" % in_XLFile
    sql = (
        "SELECT * FROM [%s].[%s%s] %s"
        % (strXLFile, strFieldsSheet, rangeFlds, whereAll)
        )
    rs.Open(sql, conn)
    if not rs:
        gp.AddMessage("No rs object from Open!")
    nrows = 0
    flds = []
    strExtraInfo = "; %s; getting Fields.Count %s" % (strGender, repr(rs))
    nf = rs.Fields.Count
    strExtraInfo = "; %s; getting keys %s" % (strGender, repr(rs.Fields))
    keys = [rs.Fields.Item(i).Name for i in xrange(nf)]
    sheets = {} # keep track of sheets while we loop
    ageFactor = 1

    strExtraInfo = "; %s; getting Fields.Count %s" % (strGender, repr(rs))
    while not rs.EOF:
        nrows += 1
        strExtraInfo = "; %s; getting values; row %d\n%s" % (strGender,
            nrows, repr(rs.Fields))
        values = [rs.Fields.Item(i).Value for i in xrange(nf)]
        row = dict(zip(keys, values))

```

```

    if strAgeGroup != "all":
        if strAgeGroup == "open": reqMaxAge=max(maxOpenAge,row['MinAge']+5)
# set the ageFactor to 1 for coincident ranges and the last open age group
        if (reqMinAge==row['MinAge']
            and reqMaxAge==row['MaxAge']) or row['MaxAge'] is None:
            ageFactor = 1
        else:
            reqRange = min(reqMaxAge,row['MaxAge'])-max(reqMinAge,row['MinAge']+1)
            ageFactor = reqRange/(row['MaxAge']-row['MinAge']+1)

row['FacAge'] = ageFactor
iMinAge = int(round(row['MinAge'],0))
row['CYGrp'] = lutCYGrp[iMinAge] if iMinAge <= MaxCYlut else MaxCYGrp
sheets[row['Sheet']] = 1 # flag that we need this sheet
whereArea = " %s AND [Type]='AREA' AND [Sheet]='%s'" % (whereBase, row['Sheet'])

strExtraInfo = "; %s; win32com; row %d" % (strGender,nrows)
rs2 = win32com.client.Dispatch("ADODB.Recordset")
sql2 = (
    "SELECT * FROM [%s].[%s$%s] %s"
    % (strXLFile, strFieldsSheet, rangeFlds, whereArea)
)
rs2.Open(sql2, conn)
if not rs2:
    gp.AddMessage("No rs2 object from Open!")
nf2 = rs2.Fields.Count
keys2 = [rs2.Fields.Item(i).Name for i in xrange(nf2)]
values2 = [rs2.Fields.Item(i).Value for i in xrange(nf2)]
row2 = dict(zip(keys,values2))
row['AreaField'] = row2['Field']
rs2.Close()
del rs2

strExtraInfo = "; %s; append; row %d" % (strGender,nrows)
flds.append(row)
strExtraInfo = "; %s; MoveNext; row %d" % (strGender,nrows)
rs.MoveNext()
rs.Close()
del rs
numFlds = nrows

strExtraInfo = "; %s; reading data" % strGender
gp.AddMessage("Reading data...")
# we now know where the data is - now to get the data
# if have more than one field then could also have more than one sheet

rangeData = ""
whereData = ""

# jet equiv of full outer join: union inner join with left and right outer joins...
# Could have multiple sheets, same fields on different sheets
# e.g. different sheets for different regions
# so loop over all (unique) sheets and combine data here rather than through sql

# the following works for any number of sheets & fields

# change sheets dictionary to contain list rather than just flag
# list will contain: area field, list of (field, agefactor) pairs
for k in sheets.keys(): sheets[k] = []
for fld in flds:
    strExtraInfo = "; %s; reading data; fld %s" % (strGender,fld)
    sheet = fld['Sheet']
    if len(sheets[sheet]) == 0 :
        sheets[sheet] = [fld['AreaField'],[(fld['Field'],fld['FacAge'],fld['CYGrp'])]]
    else: sheets[sheet][1].append((fld['Field'],fld['FacAge'],fld['CYGrp']))

for sheet, sheetData in sheets.iteritems():
    strExtraInfo = "; %s; fields; sheet %s" % (strGender,sheet)
    fldArea = sheetData[0]
    fieldList = "[%s]" % fldArea
    (flds, facts, grps) = zip(*sheetData[1])
    for fldData in flds: fieldList = "%s, [%s]" % (fieldList, fldData)

strExtraInfo = "; %s; win32com; sheet %s" % (strGender,sheet)
rs = win32com.client.Dispatch("ADODB.Recordset")
sql = (
    "SELECT %s FROM [%s].[%s$%s] %s"
    % (fieldList, strXLFile, sheet, rangeData, whereData)
)
#gp.AddMessage("using sql on excel file:\n%s" % sql)

rs.Open(sql, conn)
nf = rs.Fields.Count
keys = [rs.Fields.Item(i).Name for i in xrange(nf)]
numrows = 0
while not rs.EOF:
    numrows += 1
    values = [rs.Fields.Item(i).Value for i in xrange(nf)]
    row = dict(zip(keys,values))
    data[strGender][row[fldArea]] = [0 for cyg in range(0, MaxCYGrp+1)]

```

```

        for ff in sheetData[1]:
            (fldData, facAge, cygrp) = ff
            d = (row[fldData] if row[fldData] is not None else 0) * facAge
            data[strGender][row[fldArea]][cygrp] = (
                data[strGender][row[fldArea]][cygrp]
                if row[fldArea] in data[strGender] else 0) + d
        rs.MoveNext()
    rs.Close
    del rs

# end of gender loop

# need to add data to table...
strCodeSec = "creating table"
strExtraInfo = ""

if gp.Exists(strTmpPopTab):
    gp.Delete(strTmpPopTab)
gp.CreateTable(in_ScratchWDS, strTmpPopTabName)

gp.AddField(strTmpPopTab, strAreaField, "text", 50)
strCodeSec = "adding fields"
strExtraInfo = ""
for g in allGs:
    for cyg in range(cyGrpLo, cyGrpHi+1):
        strExtraInfo = "; %s, %s, %s" % (g, cyg, strPopField[g][cyg])
        gp.AddField(strTmpPopTab, strPopField[g][cyg], "long")

gp.AddMessage("Writing data...")
strCodeSec = "writing data"
strExtraInfo = "; insert cursor"
# create insert cursor for new table
ins_rows = gp.InsertCursor(strTmpPopTab)

# useful to swap order of keys...
# NB: allGs == data.keys() so could use either here
dKeys = set()
for k in data.keys(): dKeys = dKeys.union(data[k].keys())
data2 = dict( [k2, dict( [ k1, data[k1][k2] ] for k1 in data.keys() ) ] for k2 in dKeys )
# pop now a dictionary of allGs
for area, pop in data2.iteritems():
    strExtraInfo = "; adding data: area, pop: %s\n%s" % (area, repr(pop))
    if area:
        ins_row = ins_rows.NewRow()
        ins_row.SetValue(strAreaField, area)
        for g in allGs:
            for cyg in range(cyGrpLo, cyGrpHi+1):
                ins_row.SetValue(strPopField[g][cyg], long(round(pop[g][cyg])))

        ins_rows.InsertRow(ins_row)

strCodeSec = "copying table"
strExtraInfo = ""
gp.CopyRows_management(strTmpPopTab, strPopTab, "")

conn.Close()
del conn

### Done ###
gp.AddMessage("Done.")
strCodeSec = "End!"
strExtraInfo = ""

if blnKTO:
    strCodeSec = "exception to keep tool open!"
    exit() # keep tool open in case works and gets to here!

except:
    gp.AddMessage("Got exception in SelectPopData, section = %s %s" % (strCodeSec, strExtraInfo))
    raise
else:
    return 0 # exit errorlessly

if __name__ == '__main__':
    main(*sys.argv[1:])

```

SelectEpochData.py

```
# -----
# SelectEpochData.py
# -----

# Import system modules
import sys, string, os, math, arcgisscripting

import LatitudeFunctions as LF
import LatitudeDataNames as LDN

# Create the Geoprocessor object
gp = arcgisscripting.create(9.3)

# a few gp constants
strNKF = "NO_KEY_FIELD"
strJKA = "KEEP_ALL"
strJKC = "KEEP_COMMON"
strSNS = "NEW_SELECTION"
strSSS = "SUBSET_SELECTION"

def main(*argv):
    strCodeSec = "getting parameters"
    strExtraInfo = ""
    try:

        # use unqualified field names where possible
        # save current setting first
        strDefQualFieldNames = gp.QualifiedFieldNames
        gp.QualifiedFieldNames = "UNQUALIFIED"

        # get basic x-y epoch data from access table filtered on gender
        # 'fix' and get culmulative months
        # select according to required NZ residence criterion
        # - e.g. all prior to onset, only years prior, or specific age range
        # dissolve on case id (or fk) to give multipart features
        # select on features completely within NZ area - use latitude grid/box
        # make simple table of required cases (fk's)
        # join table back to epoch data and select epochs for required cases
        # select according to required life period - must be subset of residence
        # period unless prev where know must be resident since in census db
        # aggregate latitude data by case - e.g. by weighted average of epochs
        # construct table of latitude output data and store in gdb
        # also construct histogram of years involved and summary stats

        strCodeSec = "getting parameters"
        strExtraInfo = ""

        # Script arguments...
        iParam=0
        in_EpochDB = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_EpochDB == '#':    in_EpochDB = ""
        iParam=iParam+1
        in_PreFilter = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_PreFilter == '#':    in_PreFilter = ""
        iParam=iParam+1
        in_ResPeriod = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_ResPeriod == '#':    in_ResPeriod = ""
        iParam=iParam+1
        in_GridWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_GridWDS == '#':    in_GridWDS = LDN.WDSGrid()
        iParam=iParam+1
        in_AreasWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_AreasWDS == '#':    in_AreasWDS = LDN.WDSAreas()
        iParam=iParam+1
        in_CYear = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_CYear == '#':    in_CYear = ""
        iParam=iParam+1
        in_LifePeriod = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_LifePeriod == '#':    in_LifePeriod = ""
        iParam=iParam+1
        in_Factor = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_Factor == '#':    in_Factor = ""
        iParam=iParam+1
        in_MSLatWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_MSLatWDS == '#':    in_MSLatWDS = LDN.WDSMSLat()
        iParam=iParam+1
        in_ScratchWDSn = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_ScratchWDSn == '#':    in_ScratchWDSn = LDN.WDSScratch()
        iParam=iParam+1
        in_KTO = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_KTO == '#':    in_KTO = ""
        iParam=iParam+1

        strCodeSec = "parsing parameters"
        strExtraInfo = ""
```

```

blnKTO = (in_KTO.lower() == "true")

if not gp.Exists(in_EpochDB):
    gp.AddMessage("Input Access Database file %s does not exist!" % in_EpochDB)
    return 0
blnPreFilter = (in_PreFilter > "")
if blnPreFilter and not gp.Exists(in_PreFilter):
    gp.AddMessage("Input Pre-Filter table %s does not exist!" % in_PreFilter)
    return 0
if not gp.Exists(in_Factor):
    gp.AddMessage("Input Factor table %s does not exist!" % in_Factor)
    return 0

lstScratchWDS = in_ScratchWDSn.split(";")
in_ScratchWDS = lstScratchWDS[0]
if lstScratchWDS[0]=="in_memory":
    if len(lstScratchWDS) > 1:
        in_ScratchWDS2 = lstScratchWDS[1]
    else:
        in_ScratchWDS2 = LDN.WDSScratch()
else:
    in_ScratchWDS2 = in_ScratchWDS

strResGridName = LDN.KeepLatGrid("00")
strResGrid = in_GridWDS + "/" + strResGridName
if not gp.Exists(strResGrid):
    gp.AddMessage("Input Grid feature class %s does not exist - creating..." % strResGridName)
    import CreateLatitudeGrid
    CreateLatitudeGrid.main(in_MSLatWDS, "00", "#", "#", "#", "#", "#", in_ScratchWDS,"f")
    if not gp.Exists(strResGrid):
        gp.AddMessage("Input Grid feature class %s does not exist" % strResGridName)
        gp.AddMessage("called CreateLatitudeGrid with args:\n %s"
            % repr((in_MSLatWDS, "00", "#", "#", "#", "#", "#", in_ScratchWDS,"f")))
        return 0
    else:
        gp.AddMessage("Created.")

strCodeSec = "local variables"
strExtraInfo = ""

strTab_epoch = "t_data_epoch"
strTab_prev = "t_data_prev"
strTab_uin = "t_data_uin"
strEpochTab = in_EpochDB + "/" + strTab_epoch
strPrevTab = in_EpochDB + "/" + strTab_prev
strCaseTab = in_EpochDB + "/" + strTab_uin

# input table of factors
strFactorTab = in_Factor
strFactor = gp.Describe(strFactorTab).BaseName
allFs = LDN.AllFactors(strFactor)

# table and field names intrinsic part of db and process logic
# so don't use parameters - set later

# Local variables...
NoLat = LDN.NoLat()

allGs = LDN.AllGenders() # ['a', 'f', 'm']
strCYear = LF.parseCYear(in_CYear)
# census year groups; last group open ended
# group definitions, year limits, month limits
allCYs = LDN.AllYearGroups(strCYear)
allCYLFlds = LDN.AllCYLatFieldNames(strCYear)
allCYUFlds = LDN.AllCYUrbFieldNames(strCYear)
allCYLims = LDN.AllCYYearLimits(strCYear)
allCYMLims = LDN.AllCYMthLimits(strCYear)
numCYs = [ allCYLims[i][1] - allCYLims[i][0] + 1 for i in range(0,len(allCYLims) - 1) ]
numCYs.append(None)
numCYMs = [ allCYMLims[i][1] - allCYMLims[i][0] for i in range(0,len(allCYMLims) - 1) ]
numCYMs.append(None)
lutCYGrp = []
MaxCYGrp = len(allCYLims) - 1
# lut for years to groups - not too many years
# - too many months for equivalent month lut
for i in range(0,len(allCYLims) - 1):
    lutCYGrp.extend([i] * (numCYs[i]))
lutCYGrp.extend([len(allCYLims) - 1])
MaxCYlut = len(lutCYGrp) - 1 # should = allCYLims[len(allCYLims)-1][0]
# mainly for prev
# census at 7 March 2006
CensusYear = 2006
CensusMonth = 3

# function used for both res period and life period
# need field names defined before can call this function

#####
def whereEpochPeriod(strPeriod,reqMinMonth, reqMaxMonth):
    if strPeriod=="all":
        posn = 0 # all epochs from birth to onset

```



```

        strWhere = " [%s] > 0 " % strFld_eno
    elif strPeriod == "prev":
        posn = 6 # special prevalence point
        strWhere = " [%s] = %4d " % (strFld_toy, CensusYear)
    elif strPeriod == "preb":
        posn = 1 # have just pre-birth period
        strWhere = " [%s] = 0 " % strFld_eno
    elif strPeriod == "pbage":
        posn = 2 # have pre-birth plus age period
        strWhere = " [%s] = 0 OR [%s] < %d " % (strFld_eno, strFld_sc, reqMaxMonth)
    else:
        # have age period or period prior to onset
        posn = 3
        if strPeriod == "onset":
            posn = 4 # have period prior to onset
            strFs = strFld_src
            strFe = strFld_erc
        else: # strPeriod == "age"
            posn = 5 # have age period
            strFs = strFld_sc
            strFe = strFld_ec
        strWhere = " [%s] > 0 AND [%s] < %d AND [%s] > %d " \
            % (strFld_eno, strFs, reqMaxMonth, strFe, reqMinMonth)

    return strWhere
#####

# don't need resYears but use reqYears for census groups and year histogram
(strRPer, strRPeriod, resYears, (resMinMth, resMaxMth)) = LF.parsePeriod(in_ResPeriod)
(strLPer, strLPeriod, (reqMinYrs, reqMaxYrs),
 (reqMinMth, reqMaxMth)) = LF.parsePeriod(in_LifePeriod)
# note can only have blnRPrev if also have blnPrev
blnResPrev = (strRPeriod == "prev")
blnPrev = (strLPeriod == "prev")

# no need to check if looking at prevalence since no requirement for subset
# (fn just returns ok) - also must be resident or wouldn't be in DB
if not blnPrev:
    if not LF.checkPeriods(strRPer, strRPeriod, strLPer, strLPeriod):
        gp.AddMessage("Life Period should be a subset of residence Period\n" +
            "got Life Period, Residence Period = %s, %s" % (strLPer, strRPer))
        exit(0)

# lots of intermediate feature classes, layers, tables and views!...

strCodeSec = "setting up table and view names"
strExtraInfo = ""

# Feature Class or Table names

# to keep around
strEpochCulmsName = LDN.KeepEpochCulms()
strSEpochFCName = LDN.KeepEpochs(strRPer)
strSLatsTabName = LDN.KeepEpochLats(strRPer, strLPer)
strYHistTabName = LDN.KeepEpochHist(strRPer, strLPer)
strYStatsTabName = LDN.KeepEpochStats(strRPer, strLPer)

strUrbanFCName = LDN.KeepStatUrban(strCYear)

# intermediate
strTmpUinTabName = "tmpUin"
strTmpCaseTabName = "tmpCase"
strTmpEpochTabName = "tmpEpoch"
strTmpECTabName = "tmpCulms"
strTmpEpochFCName = "tmpEpochFC"
strTmpREpochFCName = "tmpREpochFC"
strTmpRUinsFCName = "tmpRUinsFC"
strTmpLatTabName = "tmpLat"
strTmpHistTabName = "tmpHist"
strTmpYearsTabName = "tmpYears"
strTmpStatsTabName = "tmpStats"
strMemYearsTabName = "tmpYears"
strMemStatsTabName = "tmpStats"

# ... and qualified with Feature Dataset or Workspace
strMemoryWDS = "in_memory"

strEpochCulms = in_MSLatWDS + "/" + strEpochCulmsName
strSEpochFC = in_MSLatWDS + "/" + strSEpochFCName
strSLatsTab = in_MSLatWDS + "/" + strSLatsTabName
strYHistTab = in_MSLatWDS + "/" + strYHistTabName
strYStatsTab = in_MSLatWDS + "/" + strYStatsTabName

strUrbanFC = in_AreasWDS + "/" + strUrbanFCName

strTmpUinTab = in_ScratchWDS + "/" + strTmpUinTabName
strTmpCaseTab = in_ScratchWDS + "/" + strTmpCaseTabName
strTmpEpochTab = in_ScratchWDS + "/" + strTmpEpochTabName
strTmpECTab = in_ScratchWDS2 + "/" + strTmpECTabName
strTmpEpochFC = in_ScratchWDS + "/" + strTmpEpochFCName
strTmpREpochFC = in_ScratchWDS + "/" + strTmpREpochFCName
strTmpRUinsFC = in_ScratchWDS + "/" + strTmpRUinsFCName

```

```

strTmpLatTab = in_ScratchWDS2 + "/" + strTmpLatTabName
strTmpHistTab = in_ScratchWDS + "/" + strTmpHistTabName
strTmpYearsTab = in_ScratchWDS2 + "/" + strTmpYearsTabName
strTmpStatsTab = in_ScratchWDS + "/" + strTmpStatsTabName
strMemYearsTab = strMemoryWDS + "/" + strMemYearsTabName
strMemStatsTab = strMemoryWDS + "/" + strMemStatsTabName

# Layer or Table View names
strTmpCaseView = "tmpCase_View"
strTmpFacView = "tmpFac_View"
strTmpEpochView = "tmpEpoch_View"
strTmpEpochLayer = "tmpEpoch_Layer"
strTmpREpochLayer = "tmpREpoch_Layer"
strTmpRUinsView = "tmpRUins_View"
strTmpUrbanLayer = "tmpUrban_Layer"

if blnPrev:
    strEpochPrevsName = LDN.KeepEpochPrevs()
    strTmpPrvTabName = "tmpPrv"
    strTmpPrevTabName = "tmpPrevs"
    strTmpPrevView = "tmpPrev_View"
    strTmpPrvTab = in_ScratchWDS + "/" + strTmpPrvTabName
    strTmpPrevTab = in_ScratchWDS + "/" + strTmpPrevTabName
    strEpochPrevs = in_MSLatWDS + "/" + strEpochPrevsName
if blnPreFilter:
    strPreFTab = in_PreFilter
    strTmpPrefView = "tmpPref_View"
    strFld_pref = "fk_uin_pref"

# other model variables...

# fields already in the tables
strFld_idu = "id_uin" # taken over by OBJECTID
strFld_uin = "uin_id" # copy of orig "id_uin"
strFld_fem = "b_gender_F"
strFld_bm = "i_birth_m"
strFld_by = "i_birth_y"
strFld_om = "i_onset_m"
strFld_oy = "i_onset_y"
strFld_ide = "id_epoch" # taken over by OBJECTID
strFld_epo = "epoch_id" # copy of orig "id_epoch"
strFld_fku = "fk_uin"
strFld_eno = "i_epoch_num"
strFld_tom = "i_to_m"
strFld_toy = "i_to_y"
strFld_em = "i_epoch_m"
strFld_lon = "f_loc_long"
strFld_lat = "f_loc_lat"
strFld_idp = "id_prev" # taken over by OBJECTID
strFld_prv = "prev_id" # copy of orig "id_prev"

strFld_obj = "OBJECTID"
strFld_fkf = LDN.FieldFacUin()
strFld_fac = LDN.FieldFactor()

# fields to add before we join
strFld_X = "X" # copy of long but dXnull if long is Null
strFld_Y = "Y" # copy of lat but dYnull if lat is Null
strFld_sc = "i_sculm_m" # normal culm months at start of epoch
strFld_ec = "i_eculm_m" # normal culm months at end of epoch
strFld_src = "i_srculm_m" # reverse culm months at start of epoch
strFld_erc = "i_erculm_m" # reverse culm months at end of epoch

# others created
# cy and cyurb fields defined above with other cy stuff
strFld_gen = LDN.FieldGender()
strFld_eff = LDN.FieldLatEffect()
strFld_urb = LDN.FieldUrban()
strFld_yr = LDN.FieldYear()
#strFld_nyr = {}
#for g in allGs: strFld_nyr[g] = LDN.FieldNumYears(g)
strFld_nyr = dict([(g, dict([(f, LDN.FieldNumYears(g+f))
                            for f in allFs])) for g in allGs])

strFld_yst_cnt = LDN.FieldStatsCount(strFld_yr)
strFld_yst_sum = LDN.FieldStatsSum(strFld_yr)
strFld_yst_min = LDN.FieldStatsMin(strFld_yr)
strFld_yst_max = LDN.FieldStatsMax(strFld_yr)
strFld_yst_mean = LDN.FieldStatsMean(strFld_yr)
strFld_yst_std = LDN.FieldStatsStdDev(strFld_yr)
strFld_yst_kurt = LDN.FieldStatsKurt(strFld_yr)
strFld_yst_skew = LDN.FieldStatsSkew(strFld_yr)

# some epochs have no position set
# when bring in as XY Event Layer these epochs are left out
# need to assign them position outside area of interest, e.g. 0,0 here
# but need to keep track of which epochs these are, just in case
# so copy table to tmp table and add new X,Y fields
# set X,Y to 0,0 but leave original fields as are
dXnull = 0.0
dYnull = 0.0

```

```

# prev...
# if looking at prevalence then need two lots of data
# 1 - prev data to get lats for prevalence at census point
# also, if not blnResPrev then
# 2 - epochs as normal to get uins for residency during required period
# so use normal or use copy of prev data

# shouldn't use arcgis to modify Access tables
# if use odc connection rather than access db directly
# will get qualified looking, mangled field names
# - qualified then '.' changed to '_'!
# hence direct use of access db (also *much* quicker!)
#
# copy tables to tmp tables then work from those
# NB: if do straight CopyRows then loose pk field!
# - assumes access pk field is OBJECTID field so creates
#   new OBJECTID field with orig field name as alias
#   then omits orig field in copy
#
# can use MakeQueryTable as part of copy table...
# add extra term in field list (still need orig pk)
# - field list doesn't have to be fields so use "fld+0 alias"
# - calculated field won't be treated as OBJECTID
# but make sure new name is unique
#
# could join through query but less control over qualification
# of field names so just do simple filtering and join later

# make sure we have some sort of onset date
# set to empty string if not erquired
  whereOnset = " %s.%s is not Null " % (strTab_uin, strFld_oy)

  strCodeSec = "getting data"
  strExtraInfo = ""

# leave gender til later
  whereGender = ""
  strWhere = whereOnset + whereGender

  gp.AddMessage("Reading Case data...")

# start with the new calculated field to fudge pk field
  strFieldMap = "%s.%s+0 %s" % (strTab_uin, strFld_idu, strFld_uin)
# add the other fields we need - use alias to remove extra qualification
  for fld in (strFld_idu, strFld_fem, strFld_bm, strFld_by, strFld_om, strFld_oy):
    strFieldMap += ";%s.%s %s" % (strTab_uin, fld, fld)

  gp.MakeQueryTable_management(strCaseTab, strTmpCaseView,
                              strNKF, "#", strFieldMap, strWhere)
  gp.AddMessage("Reading Factor data...")

# factors in gdb table so more straightforward
  gp.MakeQueryTable_management(strFactorTab, strTmpFacView,
                              "USE_KEY_FIELDS", strFld_obj, ";" .join([strFld_obj, strFld_fkf, strFld_fac]))
# Join to get case factors
  gp.AddJoin_management(
    strTmpCaseView, strFld_uin, strTmpFacView, strFld_fkf, strJKC)
# if need to pre-filter do it here
# need to check field names not mangled!
  if blnPreFilter:
    gp.AddMessage("Reading Pre-Filter data...")

    if gp.Exists(strTmpUinTab): gp.Delete(strTmpUinTab)
    gp.CopyRows_management(strTmpCaseView, strTmpUinTab)

    strFieldMap = ""
    gp.MakeTableView_management(
      strTmpUinTab, strTmpCaseView, "", in_ScratchWDS, strFieldMap)

    gp.MakeQueryTable_management(strPreFTab, strTmpPrefView)
# Join to get pre-filtered cases
    gp.AddJoin_management(
      strTmpCaseView, strFld_uin, strTmpPrefView, strFld_pref, strJKC)

    if gp.Exists(strTmpCaseTab): gp.Delete(strTmpCaseTab)
    gp.CopyRows_management(strTmpCaseView, strTmpCaseTab)

    strFieldMap = ""
    gp.MakeTableView_management(
      strTmpCaseTab, strTmpCaseView, "", in_ScratchWDS, strFieldMap)

#ned#[... normal epoch data
# - don't need if only using prev for res
# - need if not prev or other res for prev
  if not blnResPrev:
    gp.AddMessage("Reading Epoch data...")

    whereEpoch = ""
# start with the new calculated field to fudge pk field
    strFieldMap = "%s.%s+0 %s" % (strTab_epoch, strFld_ide, strFld_epo)

```

```

# add the other fields we need - use alias to remove extra qualification
  for fld in (strFld_ide, strFld_fku, strFld_eno,
             strFld_tom, strFld_toy, strFld_em, strFld_lat, strFld_lon):
    strFieldMap += ";%.%s %s" % (strTab_epoch, fld, fld)
gp.MakeQueryTable_management(strEpochTab, strTmpEpochView ,
                             strNKF, "#", strFieldMap, whereEpoch)

    if gp.Exists(strTmpEpochTab): gp.Delete(strTmpEpochTab)
gp.CopyRows_management(strTmpEpochView , strTmpEpochTab)
gp.AddField(strTmpEpochTab, strFld_X, "float")
gp.AddField(strTmpEpochTab, strFld_Y, "float")

# since will need both start and end of epoch periods record both here
gp.AddField(strTmpEpochTab, strFld_sc, "long")
gp.AddField(strTmpEpochTab, strFld_ec, "long")
gp.AddField(strTmpEpochTab, strFld_src, "long")
gp.AddField(strTmpEpochTab, strFld_erc, "long")

gp.AddMessage("Combining data...")

# need to truncate epochs at onset - onset info in uin table
# then cycle through setting the culm_m entries
# need views to join

    strFieldMap = ""
gp.MakeTableView_management(
    strTmpEpochTab, strTmpEpochView, "", in_ScratchWDS, strFieldMap)

# Join to get epochs for required cases
gp.AddJoin_management(
    strTmpEpochView, strFld_fku, strTmpCaseView, strFld_uin, strJKC)

    #gp.AddMessage("Done join")

# can't create update cursor for joined table view
# and can't query on right part of join
# so yet another table...
gp.CopyRows_management(strTmpEpochView, strTmpECTab)
# tidy up a bit... though don't use strTmpEpochView again
gp.RemoveJoin_management(strTmpEpochView, strTmpCaseTabName)
del strTmpEpochView
#ned#[...]
#prv#[...]
if blnPrev:
    gp.AddMessage("Reading Prev data...")

    wherePrev = ""
# start with the new calculated field to fudge pk field
    strFieldMap = "%. %s+0 %s" % (strTab_prev, strFld_idp, strFld_prv)
# add the other fields we need - use alias to remove extra qualification
    for fld in (strFld_idp, strFld_fku, strFld_lat, strFld_lon):
        strFieldMap += ";%.%s %s" % (strTab_prev, fld, fld)
    strFieldMap += ";%. %s+0.0 %s" % (strTab_prev, strFld_lon, strFld_X)
    strFieldMap += ";%. %s+0.0 %s" % (strTab_prev, strFld_lat, strFld_Y)
    strFieldMap += ";99 %s" % (strFld_eno)
    strFieldMap += ";%d %s" % (CensusMonth, strFld_tom)
    strFieldMap += ";%d %s" % (CensusYear, strFld_toy)
    strFieldMap += ";%d %s" % (strFld_em)
gp.MakeQueryTable_management(strPrevTab, strTmpPrevView ,
                             strNKF, "#", strFieldMap, wherePrev)

    if gp.Exists(strTmpPrvTab): gp.Delete(strTmpPrvTab)
gp.CopyRows_management(strTmpPrevView , strTmpPrvTab)

gp.AddMessage("Combining data...")

    strFieldMap = ""
gp.MakeTableView_management(
    strTmpPrvTab, strTmpPrevView, "", in_ScratchWDS, strFieldMap)

# Join to get prevs for required cases
gp.AddJoin_management(
    strTmpPrevView, strFld_fku, strTmpCaseView, strFld_uin, strJKC)

    #gp.AddMessage("Done join")

gp.CopyRows_management(strTmpPrevView, strTmpPrevTab)
# tidy up a bit... though don't use strTmpPrevView again
gp.RemoveJoin_management(strTmpPrevView, strTmpCaseTabName)
del strTmpPrevView
#prv#[...]
del strTmpCaseTabName

# use strTmpECTab instead of strEpochCulms then copy at end
# will avoid file of incomplete data lying around in case of error

#clm#[... fixing epoch culms
# - don't need if only using prev for res
# - need if not prev or other res for prev
if not blnResPrev:
    strCodeSec = "fixing epoch culms"

```

```

strExtraInfo = ""

gp.AddMessage("Calculating data...")

whereOnset = ("((%s - %s) * 12 + switch(%s is Null, 0, True, %s) - " +
    "switch(%s is Null, 1, %s = 0, 1, True, %s)) < %s") % \
    (strFld_toy, strFld_oy, strFld_tom, strFld_tom,
    strFld_om, strFld_om, strFld_om, strFld_em)
strSortBy = "%s A; %s A" % (strFld_fku, strFld_eno)
strExtraInfo = "%s\nstrTmpECTab = %s\nwhereOnset = %s\nstrSortBy = %s" % (
    "getting update cursor", strTmpECTab, whereOnset, strSortBy)
rows = gp.UpdateCursor(strTmpECTab, whereOnset, "", "", strSortBy)

strExtraInfo = "getting first row"
row = rows.next()
if row is None:
    gp.AddMessage("row is None!")
    gp.AddMessage("table: %s\nwhereOnset: %s\nstrSortBy: %s"
        % (strEpochCulms, whereOnset, strSortBy))
strExtraInfo = "getting first row data"
nextUin = row.GetValue(strFld_fku)
currentUin = nextUin

strExtraInfo = "Iterating through rows of cursor to calculate culms"

# Iterate through the rows in the cursor
while row:
    # iterate over all rows - this bit done once per uin
    try:
        strExtraInfo = "getting the initial data"
        posn = 0
        iom = row.GetValue(strFld_om)
        if iom == "" or iom == 0 or iom is None: iom = 1
        iMonset = (row.GetValue(strFld_oy) - row.GetValue(strFld_by)) \
            * 12 + iom - row.GetValue(strFld_bm)
        iMtot = 0

        # while we're here, do the long lats
        strExtraInfo = "doing the initial long lats"
        x = row.GetValue(strFld_lon)
        y = row.GetValue(strFld_lat)
        row.SetValue(strFld_X, (dXnull if (x == "" or x is None) else x))
        row.SetValue(strFld_Y, (dYnull if (y == "" or y is None) else y))

        rows.UpdateRow(row)
        row = rows.next()

    except:
        gp.AddMessage("exception in first part of iteration, posn = %s; " \
            + "currentUin = %s; x = %s" % (repr(posn), repr(currentUin), repr(x)))
        raise

    # cursor sorted by uin (fk) then epoch num so first row per uin
    # will be epoch 0 - i.e. pre-birth - copied then changed from
    # epoch 1 so should always be a next row with same uin
    # leave culm values as default null from AddField - can't set to
    # null/None here using SetValue unless 9.3.1 SP1 so assume can't!
    # just set long lats and go to next row
    try:
        while row and nextUin and currentUin == nextUin:
            # iterate over all rows with the same uin
            # if already up to onset #(or prev)# then discard epoch
            posn = 0
            if iMtot == iMonset:
                strExtraInfo = "deleting the extra row data"
                rows.DeleteRow(row)
            else:
                strExtraInfo = "getting the row data"
                # determine culm_m entries
                row.SetValue(strFld_sc, iMtot)
                row.SetValue(strFld_erc, iMonset-iMtot)
                iMtot += row.GetValue(strFld_em)
                if iMtot > iMonset:
                    row.SetValue(strFld_em,
                        row.GetValue(strFld_em) + iMonset - iMtot)
                    iMtot = iMonset

                row.SetValue(strFld_ec, iMtot)
                row.SetValue(strFld_src, iMonset-iMtot)

                # while we're here, do the long lats
                strExtraInfo = "doing the row long lats"
                x = row.GetValue(strFld_lon)
                y = row.GetValue(strFld_lat)
                row.SetValue(strFld_X, (dXnull if (x == "" or x is None) else x))
                row.SetValue(strFld_Y, (dYnull if (y == "" or y is None) else y))

                rows.UpdateRow(row)

            row = rows.next()
        if row: nextUin = row.GetValue(strFld_fku)

```

```

        currentUin = nextUin
    except:
        gp.AddMessage(("exception in second part of iteration, posn = %s; "
            + "currentUin = %s") % (repr(posn),repr(currentUin)))
        raise

    del row
    del rows
#clm#...]

strExtraInfo = "copying rows"
if not blnResPrev: gp.CopyRows_management(strTmpECTab, strEpochCulms)

if blnPrev: gp.CopyRows_management(strTmpPrevTab, strEpochPrevs)

strCodeSec = "determining residency"
strExtraInfo = ""

gp.AddMessage("Creating features...")

# epochs now 'fixed' and have culm values...
# strEpochCulms has correct epochs and culms
# but need to use strEpochPrevs if prev res period
if blnResPrev:
    strEpochRes = strEpochPrevs
else:
    strEpochRes = strEpochCulms

# need spatial reference - take from grid
spRefDeg = gp.Describe(strResGrid).SpatialReference

# Create layer from XY epoch data
# - specify spatial reference else problems when look at urban
gp.MakeXYEventLayer_management(strEpochRes, "X", "Y", strTmpEpochLayer, spRefDeg)

gp.AddMessage("Determining residency...")

# Select epochs based on residence period

# parsing of ResPeriod now done at beginning for coding
# but still need where clause
whereRPerEpoch = whereEpochPeriod(strRPeriod, resMinMth, resMaxMth)

strSelTypeByAttrib = strSNS
gp.SelectLayerByAttribute_management(strTmpEpochLayer, strSelTypeByAttrib, whereRPerEpoch)

# Need a feature class of selected features rather than layer
gp.CopyFeatures_management(strTmpEpochLayer, strTmpEpochFC, "", "0", "0", "0")
# prev - last use of residency strTmpEpochLayer
# re-do based on strEpochPrevs instead of strEpochCulms
if blnPrev:
    gp.MakeXYEventLayer_management(strEpochPrevs, "X", "Y", strTmpEpochLayer, spRefDeg)
    wherePrevEpoch = whereEpochPeriod(strLPeriod, "", "")
    gp.SelectLayerByAttribute_management(strTmpEpochLayer,
        strSelTypeByAttrib, wherePrevEpoch)

# Dissolve features to give one multi-part feature per uin
gp.Dissolve_management(strTmpEpochFC, strTmpREpochFC,
    strFld_fku, "", "MULTI_PART", "DISSOLVE_LINES")

# Now need a Layer... rename uin field since want to join back
strResCaseField = strFld_fku + "_res"
strFieldInfo = "%s %s VISIBLE NONE" % (strFld_fku, strResCaseField)
gp.MakeFeatureLayer_management(strTmpREpochFC, strTmpREpochLayer, "", "", strFieldInfo)

# Find features according to residence criterion
strWhereByLoc = "COMPLETELY_WITHIN"
gp.SelectLayerByLocation_management(strTmpREpochLayer, strWhereByLoc, strResGrid, "", strSNS)

# Need a feature class of the selected features rather than layer
gp.CopyFeatures_management(strTmpREpochLayer, strTmpRUinsFC, "", "0", "0", "0")

# And a Table View of the uin list
gp.MakeTableView_management(strTmpRUinsFC, strTmpRUinsView, "", in_ScratchWDS, "")

# check:
# if join KEEP_COMMON then CopyRows, which rows get copied?
#
# Join uin table back onto epoch layer
#
gp.AddJoin_management(strTmpEpochLayer, strFld_fku, strTmpRUinsView, strResCaseField, strJKA)

# Select epochs for required uins
strSelTypeByAttrib = strSNS
strWhereByAttrib = "%s is not Null" % strResCaseField
gp.SelectLayerByAttribute_management(strTmpEpochLayer, strSelTypeByAttrib, strWhereByAttrib)

# Process: Remove Join... Join name is underlying FC not View name...
gp.RemoveJoin_management(strTmpEpochLayer, strTmpRUinsFCName)

```

```

# flag urban/rural
gp.MakeFeatureLayer_management(strUrbanFC, strTmpUrbanLayer)

# could use spatial join but very slow
# also, field mappings not straightforward to encode
# gp.SpatialJoin_analysis(strTmpEpochLayer, strTmpUrbanLayer, strTmpUrbanFC,
# "JOIN_ONE_TO_ONE", "KEEP_ALL", strFieldMap, "INTERSECTS", )
# alternative...

# add new field and set default to "rural" for selected epochs
gp.AddField(strTmpEpochLayer, strFld_urb, "text", 0, 0, 5)
gp.CalculateField_management(strTmpEpochLayer, strFld_urb, "\"rural\"", "VB", "")

# select by urban location from current selection
# - if omit units assumes units of spatial ref, i.e. degrees!
strWhereByLoc = "INTERSECT"
sBuffer = "" # "1000 Meters"
gp.SelectLayerByLocation_management(strTmpEpochLayer,
strWhereByLoc, strTmpUrbanLayer, sBuffer, strSSS)

# set remaining selected epochs to urban
gp.CalculateField_management(strTmpEpochLayer, strFld_urb, "\"urban\"", "VB", "")

# re-select required epochs - should be all with urban set to something
strSelTypeByAttrib = strSNS
strWhereByAttrib = "%s is not Null" % strFld_urb
gp.SelectLayerByAttribute_management(strTmpEpochLayer, strSelTypeByAttrib, strWhereByAttrib)

# Need a feature class of the selected features rather than layer
gp.CopyFeatures_management(strTmpEpochLayer, strSEpochFC, "", "0", "0", "0")

# reset qualified field names setting
gp.QualifiedFieldNames = strDefQualFieldNames

strCodeSec = "filtering period"
strExtraInfo = ""

gp.AddMessage("Filtering period...")

# now have FC of epochs of cases we're interested in
# still need to find intersection of these with life-period
# unless prev where already done when generating prev epoch tab
# where clause therefore superfluous but leave for simplicity

# use where clause of search cursor
# as well as standard epoch data, have 0 epoch entries for pre-birth
# epochs truncated at onset to establish i_rculm_m - reverse count
# data has start date and duration for each epoch and date of onset

# parsing of LifePeriod now done at beginning for coding
# but still need where clause
whereLPerEpoch = whereEpochPeriod(strLPeriod, reqMinMth, reqMaxMth)

strSortBy = "%s A" % (strFld_fku)
rows = gp.searchcursor(strSEpochFC, whereLPerEpoch, "", "", strSortBy)

# also need tables for data

# need to keep track of actual age and bin into census groups
# reqMinYrs & reqMaxYrs are extremes of life period to be considered
# if all, prevalence or prior to onset then need all census groups
# else looking at age period (with or without pre-birth)
# and period will intersect fixed subset of census groups

if strLPeriod in ("onset", "prev"):
    iMinCYrGrp = 0
    iMaxCYrGrp = MaxCYGrp
else:
    iMinCYrGrp = lutCYGrp[min(reqMinYrs, MaxCYlut)]
    iMaxCYrGrp = lutCYGrp[min(reqMaxYrs, MaxCYlut)]

if gp.Exists(strTmpLatTab): gp.Delete(strTmpLatTab)
gp.CreateTable(in_ScratchWDS2, strTmpLatTabName)

gp.AddField(strTmpLatTab, strFld_fku, "long")
gp.AddField(strTmpLatTab, strFld_gen, "text", 0, 0, 1)
gp.AddField(strTmpLatTab, strFld_fac, "text", 0, 0, 1) # note length
for i in range(iMinCYrGrp, iMaxCYrGrp + 1):
    gp.AddField(strTmpLatTab, allCYLFlds[i], "float")
    gp.AddField(strTmpLatTab, allCYUFlds[i], "float")

# create insert cursor for new table
ins_rows = gp.InsertCursor(strTmpLatTab)

# can't have second insert cursor on same WDS?...
# use in_memory for one
if gp.Exists(strMemYearsTab): gp.Delete(strMemYearsTab)
gp.CreateTable(strMemoryWDS, strMemYearsTabName)
try:
    gp.AddField(strMemYearsTab, strFld_gen, "text", 0, 0, 1)
    gp.AddField(strMemYearsTab, strFld_fac, "text", 0, 0, 1) # note length

```

```

gp.AddField(strMemYearsTab,strFld_yr,"long")
except:
gp.AddMessage("Got exception adding field %s to %s" % (strFld_yr, strMemYearsTab))
raise

# create insert cursor for new table
try:
ins_yrows = gp.InsertCursor(strMemYearsTab)
except:
gp.AddMessage("Got exception creating cursor for %s" % strMemYearsTab)
raise

row = rows.next()
nextUin = row.GetValue(strFld_fku)
currentUin = nextUin
# in access b_gender_F is boolean
# appears as 1 for F; stats give it as 1; comparison tests give it as -1 !!!
# so check "<> 0" or "==" 0" rather than T/F or 1/0 etc
nextGen = "m" if row.GetValue(strFld_fem) == 0 else "f"
currentGen = nextGen
nextFac = row.GetValue(strFld_fac)
currentFac = nextFac

min_hist_year = 1900
max_hist_year = CensusYear # 2006
numBins = max_hist_year - min_hist_year + 1
intYear = [yr for yr in range(min_hist_year, max_hist_year + 1)]
intCYear = dict([(y, dict([(x, [0] * numBins) for x in allFs])) for y in allGs])
# keep track of min and max as we go through - first set extremes
min_h_year = dict([(y, dict([(x, max_hist_year) for x in allFs])) for y in allGs])
max_h_year = dict([(y, dict([(x, min_hist_year) for x in allFs])) for y in allGs])

strCodeSec = "getting latitudes"
strExtraInfo = ""

gp.AddMessage("Calculating latitudes...")

# Iterate through the rows in the cursor
while row:
strExtraInfo = "next uin row"

# now don't just need number of months but also which census group
# epochs and census groups in order but limits won't coincide
# use allCYMLims for month limits of census groups
# note: not just multiple of year limits: e.g. Y 0-4 -> M 0-60
dLat = [0.0] * (MaxCYGrp + 1)
dUrb = [0.0] * (MaxCYGrp + 1)
iMtot = [0] * (MaxCYGrp + 1)

strExtraInfo = "case years"
# keep track of case years for year histogram, and poss prev
# for prev, need age in years at census for year group
(case_by, case_oy) = (row.GetValue(strFld_by), row.GetValue(strFld_oy))
if strLPeriod == "prev":
prevAge = CensusYear - case_by
case_bm = row.GetValue(strFld_bm)
if case_bm > CensusMonth: prevAge -= 1
if strLPeriod == "onset":
# onset months = (sc + erc) = (src + ec)
iMthsOnset = row.GetValue(strFld_ec) + row.GetValue(strFld_src)

while row and nextUin and currentUin == nextUin:
try:
strExtraInfo = "next epoch row"

# add month weighted latitude
# 'chop' epoch if over spills req period limits
# if age or prior to onset period check both ends of epoch
# if c-age period just check end of epoch
# if c-c no check

dum = 1.0 if row.GetValue(strFld_urb) == "urban" else 0.0
vLat = row.GetValue(strFld_lat)

# overlap of epoch with required period
# subject to 'special' periods
#
# epoch months
# if onset also need reverse culms
# overlap not trivial since count back from onset,
# census groups count forward from birth
#
##### still to check works for some non- age options
if not (strLPeriod == "prev"):
if strLPeriod == "onset":
iEMinMth = row.GetValue(strFld_src)
iEMaxMth = row.GetValue(strFld_erc)
else:
iEMinMth = row.GetValue(strFld_sc)
iEMaxMth = row.GetValue(strFld_ec)
iREMinMth = max(iEMinMth, reqMinMth)

```



```

        iREMaxMth = min(iEMaxMth, reqMaxMth)

    if strLPeriod == "onset":
        iREMinYr = (iMthsOnset - iREMaxMth) // 12
        iREMaxYr = (iMthsOnset - iREMinMth - 1) // 12
    elif strLPeriod == "prev":
        iREMinYr = prevAge
        iREMaxYr = prevAge
    else:
        iREMinYr = iREMinMth // 12
        iREMaxYr = (iREMaxMth - 1) // 12

    iREMinYGrp = lutCYGrp[min(iREMinYr, MaxCYlut)]
    iREMaxYGrp = lutCYGrp[min(iREMaxYr, MaxCYlut)]

    for iYGrp in range(iREMinYGrp, iREMaxYGrp+1):
        strExtraInfo = "year group"

        if strLPeriod == "prev":
            iem = 1
        else:
            if strLPeriod == "onset":
                if iYGrp == MaxCYGrp:
                    iCREMinMth = iREMinMth
                else:
                    iCREMinMth = max(iREMinMth, (iMthsOnset - allCYMLims[iYGrp][1]))
                    iCREMaxMth = min(iREMaxMth, (iMthsOnset - allCYMLims[iYGrp][0]))
            else:
                iCREMinMth = max(iREMinMth, allCYMLims[iYGrp][0])
                if iYGrp == MaxCYGrp:
                    iCREMaxMth = iREMaxMth
                else:
                    iCREMaxMth = min(iREMaxMth, allCYMLims[iYGrp][1])
            iem = iCREMaxMth - iCREMinMth

        # if prev, only one epoch so shouldn't be here if already added
        if vLat is not None and ((not blnPrev) or (iMtot[iYGrp] == 0)):
            dLat[iYGrp] += vLat * iem
            iMtot[iYGrp] += iem
            dUrb[iYGrp] += dum * iem

    row = rows.next()
    if row:
        nextUin = row.GetValue(strFld_fku)
        nextGen = "m" if row.GetValue(strFld_fem) == 0 else "f"
        nextFac = row.GetValue(strFld_fac)
    except:
        gp.AddMessage("Got exception in inner row loop for uin = %d" % currentUin)
        raise

    strExtraInfo = "scale year group data"
    for i in range(iMinCYrGrp, iMaxCYrGrp + 1):
        dLat[i] = (dLat[i] / iMtot[i] if iMtot[i] != 0 else NoLat)
        dUrb[i] = (dUrb[i] / iMtot[i] if iMtot[i] != 0 else 0)

    strExtraInfo = "set fields"
    try:
        ins_row = ins_rows.NewRow()
        ins_row.SetValue(strFld_fku, currentUin)
        ins_row.SetValue(strFld_gen, currentGen)
        ins_row.SetValue(strFld_fac, currentFac)
        for i in range(iMinCYrGrp, iMaxCYrGrp + 1):
            ins_row.SetValue(allCYLFlDs[i], dLat[i])
            ins_row.SetValue(allCYUFlDs[i], dUrb[i])

        ins_rows.InsertRow(ins_row)
    except:
        gp.AddMessage("Got exception adding lat %s to tmp table for uin = %d"
            % (repr(dLat), currentUin))
        raise

    strExtraInfo = "year histogram"
    # determine year histogram separately - one calculation per case
    # prev gives trivial result - all census year - but include anyway
    if strLPeriod == "all":
        # use reqMinYrs (=0) and onset as offset to birth year
        min_year = case_by
        max_year = case_oy
    elif strLPeriod == "onset":
        # use reqMinYrs and reqMaxYrs as offset to onset year
        min_year = case_oy - reqMaxYrs
        max_year = case_oy - reqMinYrs
    elif strLPeriod == "prev":
        min_year = CensusYear
        max_year = CensusYear
    else:
        # use reqMinYrs and reqMaxYrs as offset to birth year
        min_year = case_by + reqMinYrs
        max_year = case_by + reqMaxYrs
        if strLPeriod[0]=="p": min_year -= 1
    if min_year < min_h_year[currentGen][currentFac]:

```

```

        min_h_year[currentGen][currentFac] = min_year
    if max_year > max_h_year[currentGen][currentFac]:
        max_h_year[currentGen][currentFac] = max_year
    for binYr in range (min_year - min_hist_year, max_year - min_hist_year + 1):
        for x in [currentGen, 'a']:
            for y in [currentFac, 'a']:
                intCYear[x][y][binYr] += 1
        try:
            ins_yr = ins_yrows.NewRow()
            ins_yr.SetValue(strFld_gen, currentGen)
            ins_yr.SetValue(strFld_fac, currentFac)
            ins_yr.SetValue(strFld_yr, binYr + min_hist_year)
            ins_yrows.InsertRow(ins_yr)
            del ins_yr
        except:
            gp.AddMessage("Got exception adding year %d to tmp table for bin %d; uin = %d"
                % (binYr + min_hist_year, binYr, currentUin))
            raise

    currentUin = nextUin
    currentGen = nextGen
    currentFac = nextFac

del row, ins_row
del rows, ins_rows
try:
    del ins_yrows
except:
    gp.AddMessage("Got exception deleting cursor")

strExtraInfo = "end"
restGs = set(allGs) - set(['a'])
restFs = set(allFs) - set(['a'])
for g in restGs:
    min_h_year[g]['a'] = min(min_h_year[g][f] for f in restFs)
    max_h_year[g]['a'] = max(max_h_year[g][f] for f in restFs)
for f in allFs:
    min_h_year['a'][f] = min(min_h_year[g][f] for g in restGs)
    max_h_year['a'][f] = max(max_h_year[g][f] for g in restGs)

if gp.Exists(strTmpYearsTab):    gp.Delete(strTmpYearsTab)
gp.CopyRows_management(strMemYearsTab, strTmpYearsTab, "")

strCodeSec = "getting stats"
strExtraInfo = "doing year stats"

sum_years = dict([(y, dict([(x, None) for x in allFs])) for y in allGs])
num_years = dict([(y, dict([(x, None) for x in allFs])) for y in allGs])
mean_year = dict([(y, dict([(x, None) for x in allFs])) for y in allGs])
std_dev = dict([(y, dict([(x, None) for x in allFs])) for y in allGs])
kurtosis = dict([(y, dict([(x, None) for x in allFs])) for y in allGs])
skewness = dict([(y, dict([(x, None) for x in allFs])) for y in allGs])

# stats...
# could let arc do stats on year table but want extra stuff
# can use gp statistics to compare some, though
# also compare with stats from excel...
# for sample of 7775 years from 1918 to 2000 get agreement to 10-12 sig.places!
# (apart from skew - only agrees to 4 sig. places)

gp.AddMessage("Calculating year stats...")
for g in allGs:
    for f in allFs:
        minBin = min_h_year[g][f] - min_hist_year
        maxBin = max_h_year[g][f] - min_hist_year
        for binYr in range(0, minBin): intCYear[g][f][binYr] = None
        for binYr in range(maxBin + 1, numBins): intCYear[g][f][binYr] = None
        num_y = 0
        sum_y = 0
        for binYr in range(minBin, maxBin + 1):
            num_y += intCYear[g][f][binYr]
            sum_y += intCYear[g][f][binYr] * intYear[binYr]

        num_years[g][f] = num_y
        sum_years[g][f] = sum_y

    if num_y > 0:
        fnum = float(num_y)
        mean_y = sum_y / fnum
        mean_year[g][f] = mean_y

    try:
        (sum_sqdev, sum_cudev, sum_qudev) = (0, 0, 0)
        for binYr in range(minBin, maxBin + 1):
            dev = intYear[binYr] - mean_y
            dev2 = dev * dev
            sum_sqdev += intCYear[g][f][binYr] * dev2
            sum_cudev += intCYear[g][f][binYr] * dev2 * dev
            sum_qudev += intCYear[g][f][binYr] * dev2 * dev2
    except:

```

```

gp.AddMessage("Got exception calculating deviations for bin %d" % binYr)
raise

if num_y > 1:
    var_y = sum_sqdev / (fnum - 1)
    std = math.sqrt(var_y)
    std_dev[g][f] = std
    if num_y > 3:
        nyp1 = num_y + 1
        nym1 = num_y - 1
        nym2 = num_y - 2
        nym3 = num_y - 3
        if var_y == 0:
            kurtosis[g][f] = -3
            skewness[g][f] = 0
        else:
# kurtosis, G2, as used by SPSS, Excel et al http://en.wikipedia.org/wiki/Kurtosis
            kurtosis[g][f] = (fnum * nyp1 * sum_sqdev
                / (var_y * var_y * nym1 * nym2 * nym3)
                - 3.0 * nym1 * nym1 / (nym2 * nym3))
# skewness, G1, http://en.wikipedia.org/wiki/Skewness
            skewness[g][f] = math.sqrt(fnum * nym1) * sum_sqdev
                / (var_y * std * fnum * nym2)

gp.AddMessage("Done year stats...")

# if need to compare...
#strStatsField = ""
#for stat in ("sum", "mean", "min", "max", "std", "count"):
#    strStatsField = strStatsField + "; %s %s" % (strFld_yr, stat)
#strCaseField = ""
#gp.Statistics_analysis(strMemYearsTab, strMemStatsTab, strStatsField[1:], strCaseField)
#
#fields = gp.ListFields(strMemStatsTab)
#rows = gp.searchcursor(strMemStatsTab)
#row = rows.next()
#gp.AddMessage("Stats from gp:")
#for fld in fields:
#    val = row.GetValue(fld.name)
#    gp.AddMessage("%s: %s" % (fld.name, repr(val)))
#
#del row
#del rows
#
#gp.AddMessage("\nStats from script:")
#gp.AddMessage("%s: %s" % (strFld_yst_sum, repr(sum_years)))
#gp.AddMessage("%s: %s" % (strFld_yst_cnt, repr(num_years)))
#gp.AddMessage("%s: %s" % (strFld_yst_min, repr(min_h_year)))
#gp.AddMessage("%s: %s" % (strFld_yst_max, repr(max_h_year)))
#gp.AddMessage("%s: %s" % (strFld_yst_mean, repr(mean_year)))
#gp.AddMessage("%s: %s" % (strFld_yst_std, repr(std_dev)))
#gp.AddMessage("%s: %s" % (strFld_yst_kurt, repr(kurtosis)))
#gp.AddMessage("%s: %s" % (strFld_yst_skew, repr(skewness)))

strExtraInfo = "saving year histogram"
# save histogram and summary stats
try:
    if gp.Exists(strTmpHistTab): gp.Delete(strTmpHistTab)
    gp.CreateTable(in_ScratchWDS, strTmpHistTabName)
except:
    gp.AddMessage("Got exception deleting or creating table %s" % strTmpHistTab)
    raise

try:
    gp.AddField(strTmpHistTab, strFld_yr, "long")
    for g in allGs:
        for f in allFs:
            gp.AddField(strTmpHistTab, strFld_nyr[g][f], "long")
except:
    gp.AddMessage("Got exception adding fields %s %s" % (strFld_yr, strFld_nyr))
    raise

# create insert cursor for new table
try:
    ins_rows = gp.InsertCursor(strTmpHistTab)
except:
    gp.AddMessage("Got exception getting insertcursor for %s" % strTmpHistTab)
    raise

for binYr in range(0, numBins):
    try:
        ins_row = ins_rows.NewRow()
    except:
        gp.AddMessage("Got exception getting ins_row for bin %d" % binYr)
        raise
    try:
        ins_row.SetValue(strFld_yr, binYr + min_hist_year)
    except:
        gp.AddMessage("Got exception setting ins_row %s value for bin %d to %s"
            % (strFld_yr, binYr, repr(binYr + min_hist_year)))
        raise

```

```

#         if intCYear[binYr] is not None:
# setting to 'None' should be fixed in 9.3.1 SP1
# else problem since will retain value from previous row!
    try:
        for g in allGs:
            for f in allFs:
                ins_row.SetValue(strFld_nyr[g][f], intCYear[g][f][binYr])
    except:
        gp.AddMessage("Got exception setting ins_row %s value for bin %d to %s"
                    % (strFld_nyr[g][f], binYr, repr(intCYear[g][f][binYr])))
        raise

    try:
        ins_rows.InsertRow(ins_row)
        del ins_row
    except:
        gp.AddMessage("Got exception inserting ins_row for bin %d" % binYr)
        raise

del ins_rows

strExtraInfo = "saving year stats"
if gp.Exists(strTmpStatsTab): gp.Delete(strTmpStatsTab)
gp.CreateTable(in_ScratchWDS, strTmpStatsTabName)

gp.AddField(strTmpStatsTab, strFld_gen, "text", 0, 0, 1)
gp.AddField(strTmpStatsTab, strFld_fac, "text", 0, 0, 1) # note length
gp.AddField(strTmpStatsTab, strFld_yst_cnt, "long")
gp.AddField(strTmpStatsTab, strFld_yst_sum, "long")
gp.AddField(strTmpStatsTab, strFld_yst_min, "long")
gp.AddField(strTmpStatsTab, strFld_yst_max, "long")
gp.AddField(strTmpStatsTab, strFld_yst_mean, "float")
gp.AddField(strTmpStatsTab, strFld_yst_std, "float")
gp.AddField(strTmpStatsTab, strFld_yst_kurt, "float")
gp.AddField(strTmpStatsTab, strFld_yst_skew, "float")

# create insert cursor for new table
ins_rows = gp.InsertCursor(strTmpStatsTab)
for g in allGs:
    for f in allFs:
        ins_row = ins_rows.NewRow()
        ins_row.SetValue(strFld_gen, g)
        ins_row.SetValue(strFld_fac, f)
        ins_row.SetValue(strFld_yst_cnt, num_years[g][f])
        ins_row.SetValue(strFld_yst_sum, sum_years[g][f])
        ins_row.SetValue(strFld_yst_min, min_h_year[g][f])
        ins_row.SetValue(strFld_yst_max, max_h_year[g][f])
        ins_row.SetValue(strFld_yst_mean, mean_year[g][f])
        ins_row.SetValue(strFld_yst_std, std_dev[g][f])
        ins_row.SetValue(strFld_yst_kurt, kurtosis[g][f])
        ins_row.SetValue(strFld_yst_skew, skewness[g][f])

        ins_rows.InsertRow(ins_row)

del ins_row
del ins_rows

strCodeSec = "writing data"
strExtraInfo = ""

gp.AddMessage("Writing data...")

gp.CopyRows_management(strTmpLatTab, strSLatsTab, "")
gp.CopyRows_management(strTmpHistTab, strYHistTab, "")
gp.CopyRows_management(strTmpStatsTab, strYStatsTab, "")

# reset qualified field names setting
gp.QualifiedFieldNames = strDefQualFieldNames

### Done ###
gp.AddMessage("Done.")
strCodeSec = "End!"
strExtraInfo = ""

if blnKTO:
    strCodeSec = "exception to keep tool open!"
    exit() # keep tool open in case works and gets to here!

except:
    gp.AddMessage("Got exception in SelectEpochData, section = %s %s" % (strCodeSec, strExtraInfo))
    raise
else:
    return 0 # exit errorlessly

if __name__ == '__main__':
    main(*sys.argv[1:])

```

SliceAreas.py

```
# -----
# SliceAreas.py
# -----

# Import system modules
import sys, string, os, arcgisscripting

import LatitudeFunctions as LF
import LatitudeDataNames as LDN

# Create the Geoprocessor object
gp = arcgisscripting.create(9.3)

def main(*argv):
    strCodeSec = "getting parameters"
    strExtraInfo = ""
    try:

        # Script arguments...
        iParam=0
        in_AreasWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_AreasWDS == '#':    in_AreasWDS = LDN.WDSAreas()
        iParam=iParam+1
        in_CYear = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_CYear == '#':    in_CYear = ""
        iParam=iParam+1
        in_GridWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_GridWDS == '#':    in_GridWDS = LDN.WDSGrid()
        iParam=iParam+1
        in_LatSize = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_LatSize == '#':    in_LatSize = ""
        iParam=iParam+1
        in_SliceWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_SliceWDS == '#':    in_SliceWDS = LDN.WDSSlice()
        iParam=iParam+1
        in_ScratchWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_ScratchWDS == '#':    in_ScratchWDS = LDN.WDSScratch()
        iParam=iParam+1
        in_KTO = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_KTO == '#':    in_KTO = ""
        iParam=iParam+1

        blnKTO = (in_KTO.lower() == "true")

        strCodeSec = "local variables"
        # local variables...
        strSlice = LF.parseLat(in_LatSize)
        strCYear = LF.parseCYear(in_CYear)

        # Feature Class or Table names
        strAreasName = LDN.KeepStatAreas(strCYear)
        strGridName = LDN.KeepLatGrid(strSlice)
        strSlicesName = LDN.KeepSlicedAreas(strCYear, strSlice)
        strTmpSlicesName = "tmpSlices"

        # ... and qualified with Feature Dataset or Workspace
        strAreas = in_AreasWDS + "/" + strAreasName
        strGrid = in_GridWDS + "/" + strGridName
        strSlices = in_SliceWDS + "/" + strSlicesName
        strTmpSlices = in_ScratchWDS + "/" + strTmpSlicesName

        strFld_sqkm = LDN.FieldSqKm()
        strFld_af = LDN.FieldAreaFrac()
        # assume this isn't already in the feature class - should check!
        strFld_tmp = "tmp_slice_area"

        # Set the scratch workspace
        gp.Workspace = in_ScratchWDS

        strCodeSec = "checking grid"
        strExtraInfo = ""
        if not gp.Exists(strGrid):
            gp.AddMessage("Input Grid feature class %s does not exist - creating..."%strGridName)
            import CreateLatitudeGrid
            strExtraInfo = "; creating"
            CreateLatitudeGrid.main(in_GridWDS,in_LatSize,"#", "#", "#", "#", "#", in_ScratchWDS, "f")
            if not gp.Exists(strGrid):
                gp.AddMessage("Input Grid feature class %s does not exist" % strGridName)
                gp.AddMessage("called CreateLatitudeGrid with args:\n %s"
                    % repr((in_GridWDS,in_LatSize,"#", "#", "#", "#", "#", in_ScratchWDS, "f")))
                return 0
            else:
                gp.AddMessage("Created.")

        strCodeSec = "checking stat areas"
        strExtraInfo = ""
        if not gp.Exists(strAreas):
```

```

gp.AddMessage("Statistical Areas feature class %s does not exist\n\n" +
              "Please run SelectAreas with manual input\n" % strAreasName)
return 0

strCodeSec = "checking temps"
strExtraInfo = ""
if gp.Exists(strTmpSlices): gp.Delete(strTmpSlices)
if gp.Exists(strSlices): gp.Delete(strSlices)

gp.AddMessage("Intersecting Features...")
strCodeSec = "intersecting features"
strExtraInfo = ""

# Intersect statistical areas and latitude grid
# must have areas first in feature class list so its spatial reference
# is used - ldeg lat ~ 110-112km for NZ giving 10^5 scale difference
# - so get *very* small areas if get it wrong!
sFeatureList = "%s #;%s #" % (strAreas, strGrid)
gp.Intersect_analysis(sFeatureList, strTmpSlices, "NO_FID", "", "INPUT")

gp.AddMessage("Calculating Areas...")

# fields for sliced areas
strCodeSec = "adding fields"
strExtraInfo = "; %s" % strFld_tmp
gp.AddField_management(strTmpSlices, strFld_tmp, "DOUBLE")
strExtraInfo = "; %s" % strFld_af
gp.AddField_management(strTmpSlices, strFld_af, "DOUBLE")

strCodeSec = "calculating fields"
# calculate the sliced areas - need the 'advanced' call
strExtraInfo = "; %s" % strFld_tmp
strCalc = "%s\n%s\n%s\n%s\n" % ("Dim Output as double", "Dim pArea as Iarea",
                                "Set pArea = [shape]", "Output = pArea.area/1000000.0")
gp.CalculateField_management(strTmpSlices, strFld_tmp, "Output", "VB", strCalc)

# and then the area fraction
strExtraInfo = "; %s" % strFld_af
strExpr = "[%s] / [%s]" % (strFld_tmp, strFld_sqkm)
gp.CalculateField_management(strTmpSlices, strFld_af, strExpr, "VB", "")

gp.AddMessage("Writing Features...")

strCodeSec = "deleting fields"
strExtraInfo = "; %s" % strFld_tmp
gp.DeleteField_management(strTmpSlices, strFld_tmp)
strExtraInfo = "; %s" % strFld_sqkm
gp.DeleteField_management(strTmpSlices, strFld_sqkm)

strCodeSec = "writing features"
gp.CopyFeatures_management(strTmpSlices, strSlices)

### Done ###
gp.AddMessage("Done.")
strCodeSec = "End!"
strExtraInfo = ""

if blnKTO:
    strCodeSec = "exception to keep tool open!"
    exit() # keep tool open in case works and gets to here!

except:
    gp.AddMessage("Got exception in SliceAreas, section = %s %s" % (strCodeSec, strExtraInfo))
    raise
else:
    return 0 # exit errorlessly

if __name__ == '__main__':
    main(*sys.argv[1:])

```

LatitudePops.py

```
# -----
# LatitudePops.py
# -----
# e.g.
# LatitudePops D:\GIS\tmp_CensusPops.mdb 2006 all
# D:\GIS\SlicedAreas.mdb 2
# D:\GIS\LatPops.mdb D:\Workspace\Scratch.mdb f

# Import system modules
import sys, string, os, arcgisscripting

import LatitudeFunctions as LF
import LatitudeDataNames as LDN

# Create the Geoprocessor object
gp = arcgisscripting.create(9.3)

def main(*argv):
    strCodeSec = "getting parameters"
    strExtraInfo = ""
    try:

        # Script arguments...
        iParam= 0
        in_PopWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_PopWDS == '#':    in_PopWDS = LDN.WDSPop()
        iParam=iParam+1
        in_CYear = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_CYear == '#':    in_CYear = ""
        iParam=iParam+1
        in_SliceWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_SliceWDS == '#':    in_SliceWDS = LDN.WDSSlice()
        iParam=iParam+1
        in_LatSize = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_LatSize == '#':    in_LatSize = ""
        iParam=iParam+1
        in_LatPopWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_LatPopWDS == '#':    in_LatPopWDS = LDN.WDSLatPop()
        iParam=iParam+1
        in_ScratchWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_ScratchWDS == '#':    in_ScratchWDS = LDN.WDSScratch()
        iParam=iParam+1
        in_KTO = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_KTO == '#':    in_KTO = ""
        iParam=iParam+1

        blnKTO = (in_KTO.lower() == "true")

        # Set the scratch workspace
        gp.Workspace = in_ScratchWDS

        # Local variables...
        strCodeSec = "local variables"
        strExtraInfo = ""

        in_AgeGroup = "all"

        strCYear = LF.parseCYear(in_CYear)
        (strAgePeriod, strPerd, (reqMinAge, reqMaxAge), reqMonths) = LF.parsePeriod(in_AgeGroup)
        strLatBin = LF.parseLat(in_LatSize)

        allGs = ['a', 'f', 'm']
        allUs = ['a', 'u', 'r']
        allGUs = [ g + u for g in allGs for u in allUs ]
        lutGU = dict( [ g, dict( [ u, g + u ] for u in allUs ) ] for g in allGs )

        # census year groups; last group open ended
        # group definitions, year limits, month limits
        allCYLims = LDN.AllCYearLimits(strCYear)
        numCYs = [ allCYLims[i][1] - allCYLims[i][0] + 1 for i in range(0,len(allCYLims) - 1) ]
        numCYs.append(None)
        lutCYGrp = []
        MaxCYGrp = len(allCYLims) - 1
        # lut for years to groups - not too many years
        for i in range(0,len(allCYLims) - 1):
            lutCYGrp.extend([i] * (numCYs[i]))
        lutCYGrp.extend([len(allCYLims) - 1])
        MaxCYlut = len(lutCYGrp) - 1 # should = allCYLims[len(allCYLims)-1][0]

        if strPerd=="all":
            cyGrpLo = 0
            cyGrpHi = MaxCYGrp
        else:
            # need to filter for high end open
            reqMinAge = int(reqMinAge)
            cyGrpLo = lutCYGrp[reqMinAge]
            if reqMaxAge=="":
```

```

        cyGrpHi = MaxCYGrp
    else:
        reqMaxAge = int(reqMaxAge)
        cyGrpHi = lutCYGrp[reqMaxAge]

usedCYGrps = range(cyGrpLo,cyGrpHi+1)

# feature classes and tables
strSlicesFCName = LDN.KeepSlicedAreas(strCYear, strLatBin)
strPopTabName = LDN.KeepPop(strCYear)

strTmpFCName = "tmpSlices"

# layer and view names
strTmpView = "tmpSlicesView"
strTmpPopView = "tmpPopsView"

strLatPopTabName = {}
strLatPopTab = {}
strTmpTabName = {}
strTmpTab = {}
strTmpStatsTabName = {}
strTmpStatsTab = {}
strTmpStatsView = {}
for gu in allGUs:
    strLatPopTabName[gu] = LDN.KeepLatPop(strCYear, strLatBin, gu)
    strTmpTabName[gu] = "tmpFracs_" + gu
    strTmpStatsTabName[gu] = "tmpStats_" + gu
    strTmpStatsView[gu] = "tmpStatsView_" + gu
    strLatPopTab[gu] = in_LatPopWDS + "/" + strLatPopTabName[gu]
    strTmpTab[gu] = in_ScratchWDS + "/" + strTmpTabName[gu]
    strTmpStatsTab[gu] = in_ScratchWDS + "/" + strTmpStatsTabName[gu]

strPopTab = in_PopWDS + "/" + strPopTabName
strSlicesFC = in_SliceWDS + "/" + strSlicesFCName
strTmpFC = in_ScratchWDS + "/" + strTmpFCName # "tmpSlices"

strFld_af = LDN.FieldAreaFrac() # "area_frac"
strFld_urb = LDN.FieldUrban()
strFld_sa = LDN.FieldStatArea() # "Census_Area" "stat_area"
strFld_pa = LDN.FieldSPopArea() # "Area_Census" "c_stat_area"
strFlds_stats = LDN.FieldListLats() # "Lat_N;Lat_S;Lat_C"

strFld_pop = {}
strFld_lp = {}
strFld_spo = {}
strFld_ssp = {}
strFlds_sum = {}

#####
# really need these for all cy grps we have data for, not all grps...
# setup for all then use limits when use
#####

for g in allGs:
    strFld_pop[g] = LDN.AllFieldsSamplePop(g,strCYear)
for gu in allGUs:
    strFld_pop[gu] = LDN.AllFieldsSamplePop(gu,strCYear)
    strFld_lp[gu] = LDN.AllFieldsLatPop(gu,strCYear)
    strFld_spo[gu] = LDN.AllCYGFldNames("%s_p" % gu,strCYear) # slice pop
    strFld_ssp[gu] = LDN.AllCYGFldNames("sum_%s_p" % gu,strCYear)
    fldS = strFld_spo[gu]
    fld_sum = ""
    for cyg in usedCYGrps:
        fld_sum = "%s;%s sum" % (fld_sum, fldS[cyg])
    strFlds_sum[gu] = fld_sum[1:]

strCodeSec = "checking input data"
strExtraInfo = ""

if not gp.Exists(strSlicesFC):
    gp.AddMessage("Input Sliced Stat Areas feature class %s does not exist - creating..."
        % strSlicesFCName)
# SliceAreas_latitude <AreasWDS>,<CYear>,<GridWDS>,<LatSize>,<SliceWDS>,<ScratchWDS>,<KTO>
import SliceAreas
SliceAreas.main("#", in_CYear, "#", in_LatSize, in_SliceWDS, in_ScratchWDS, "f")
if not gp.Exists(strSlicesFC):
    gp.AddMessage("Input Grid feature class %s does not exist" % strGridName)
    gp.AddMessage("called SliceAreas with args:\n %s"
        % repr("#", in_CYear, "#", in_LatSize, in_SliceWDS, in_ScratchWDS))
    return 0
else:
    gp.AddMessage("Created.")

if not gp.Exists(strPopTab):
    gp.AddMessage("Census Population Data table %s does not exist\n\n"
        + "Please run SelectPopData with manual input\n" % strPopTabName)
    return 0

if gp.Exists(strTmpFC):
    gp.Delete(strTmpFC)

```



```

gp.AddMessage("Reading Features...")
gp.CopyFeatures_management(strSlicesFC, strTmpFC)

if gp.Exists(strTmpView):
    gp.Delete(strTmpView)
gp.MakeTableView_management(strTmpFC, strTmpView)

# check field types for keep_fields - do it before we join
strFlds_keep = strFlds_stats.split(";")
dicFlds_type = dict()
fields = gp.ListFields(strTmpView)
for fld in fields:
    dicFlds_type[fld.name] = fld.type

gp.AddMessage("Reading Table...")
if gp.Exists(strTmpPopView):
    gp.Delete(strTmpPopView)
gp.MakeTableView_management(strPopTab, strTmpPopView)

gp.AddJoin_management(strTmpView, strFld_sa, strTmpPopView, strFld_pa, "KEEP_ALL")

# use tmp table rather than joined table to sanitise field names

gp.AddMessage("Disaggregating Data...")
for gu in allGUs:
    tabG = strTmpTab[gu]
    if gp.Exists(tabG): gp.Delete(tabG)
    gp.CreateTable(in_ScratchWDS, strTmpTabName[gu])

gp.AddMessage("Setting up new table...")
for gu in allGUs:
    tabG = strTmpTab[gu]
    for fld in strFlds_keep:
        gp.AddField(tabG, fld, dicFlds_type[fld])

# too many fields if add all here! (Get exception with just gu & cyg)
# need to split up processing in manner which will extend
# if/when add more factors such as ethnicity
# - need separate tmp tables for some or all factors
gp.AddField(tabG, strFld_af, "double")
fldP = strFld_pop[gu]
for cyg in usedCYGrps: gp.AddField(tabG, fldP[cyg], "long")
fldS = strFld_spo[gu]
for cyg in usedCYGrps: gp.AddField(tabG, fldS[cyg], "double")

# combine gender and urban into new pop fields
# scale by sliced area fraction
# use stats to aggregate (sum) new fields by latitude

#####
# can't have second insert cursor on same WDS...
#####
# strTmpFC is copy of original data
# strTmpView is view on strTmpFC with strTmpPopView joined
# after join, strTmpView fields qualified with strTmpFC name!
# use strTmpFCName variable (=strTmpFC) to qualify fields
# bit less confusing than using strTmpFC variable to qualify
# fields in search cursor on strTmpView
strTmpQual = strTmpFCName + "."
strTmpFldUrb = strTmpQual + strFld_urb
sc_where = {'a': ""}
sc_where['u'] = strTmpFldUrb + " = 'urban'"
sc_where['r'] = strTmpFldUrb + " = 'rural'"
strPTabQual = strPopTabName + "."
gp.AddMessage("Inserting new data...")
for g in allGs:
    fldG = strFld_pop[g]
    for ur in allUs:
        gu = lutGU[g][ur]
        fldP = strFld_pop[gu]
        # create search cursor for joined table
        # should have all records or only urban or rural
        get_rows = gp.SearchCursor(strTmpView, sc_where[ur])

        # create insert cursor for new table
        ins_rows = gp.InsertCursor(strTmpTab[gu])

        get_row = get_rows.Next()

        dbgrow = 0
        while get_row:
            ins_row = ins_rows.NewRow()

            for fld in strFlds_keep:
                ins_row.SetValue(fld, get_row.GetValue(strTmpQual + fld))

            ins_row.SetValue(strFld_af, get_row.GetValue(strTmpQual + strFld_af))
            for cyg in usedCYGrps:
                ins_row.SetValue(fldP[cyg], get_row.GetValue(strPTabQual + fldG[cyg]))

            ins_rows.InsertRow(ins_row)

```

```

        get_row = get_rows.Next()
        get_rows = None
        ins_rows = None

gp.AddMessage("Calculating Area Populations...")
# Don't need to qualify in new tmp table...

for gu in allGUs:
    tabG = strTmpTab[gu]
    fldP = strFld_pop[gu]
    fldS = strFld_spo[gu]
    for cyg in usedCYGrps:
        strCalc = "[%s] * [%s]" % (strFld_af, fldP[cyg])
        gp.CalculateField_management(tabG, fldS[cyg], strCalc, "VB", "")

gp.AddMessage("Calculating Slice Populations...")
gp.AddMessage("Re-aggregating Data...")
for gu in allGUs:
    tabS = strTmpStatsTab[gu]
    if gp.Exists(tabS): gp.Delete(tabS)
    gp.Statistics_analysis(strTmpTab[gu], tabS, strFlds_sum[gu], strFlds_stats)

    gp.DeleteField_management(tabS, "FREQUENCY")

gp.AddMessage("Filtering View fields...")
for gu in allGUs:
    tabV = strTmpStatsView[gu]
    if gp.Exists(tabV): gp.Delete(tabV)
    fldS = strFld_ssp[gu]
    fldP = strFld_lp[gu]
    strFieldInfo = ""
    for cyg in usedCYGrps:
        strFieldInfo = "%s;%s %s VISIBLE" % (strFieldInfo, fldS[cyg], fldP[cyg])
    strFieldInfo = strFieldInfo[1:]
    gp.MakeTableView_management(strTmpStatsTab[gu], tabV, "", "", strFieldInfo)

gp.AddMessage("Copying Rows from scratch...")
gp.AddMessage("Writing Data...")
for gu in allGUs:
    gp.CopyRows_management(strTmpStatsView[gu], strLatPopTab[gu])

### Done ###
gp.AddMessage("Done.")
strCodeSec = "End!"
strExtraInfo = ""

if blnKTO:
    strCodeSec = "exception to keep tool open!"
    exit() # keep tool open in case works and gets to here!

except:
    gp.AddMessage("Got exception in LatitudePops, section = %s %s"%(strCodeSec,strExtraInfo))
    raise
else:
    return 0 # exit errorlessly

if __name__ == '__main__':
    main(*sys.argv[1:])

```

ProcessEpochLats.py

```
# -----
# ProcessEpochLats.py
# -----

# e.g ProcessEpochData D:\GIS\EpochData.mdb
# prv prv D:\GIS\LatPops.mdb 2006 # all 2
# D:\GIS\StdPops.mdb\euro 2
# D:\GIS\NormData.mdb D:\Workspace\Scratch.mdb f

# Import system modules
import sys, string, os, math, arcgisscripting

import LatitudeFunctions as LF
import LatitudeDataNames as LDN

# Create the Geoprocessor object
gp = arcgisscripting.create(9.3)

def main(*argv):
    strCodeSec = "getting parameters"
    strExtraInfo = ""
    try:

        # read in Effective Lats for epochs and population bin data
        # bin Lats
        # normalise Lats by pop
        # save as table
        # (leave plotting to separate routine)

        # Script arguments...
        iParam=0

        in_EffWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_EffWDS == '#':      in_MSLatWDS = LDN.WDSMSLat()
        iParam=iParam+1
        in_Factor = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_Factor == '#':      in_Factor = ""
        iParam=iParam+1

        in_ResPeriod = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_ResPeriod == '#':    in_ResPeriod = ""
        iParam=iParam+1
        in_LifePeriod = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_LifePeriod == '#':   in_LifePeriod = ""
        iParam=iParam+1

        in_LatPopWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_LatPopWDS == '#':    in_LatPopWDS = LDN.WDSLatPop()
        iParam=iParam+1
        in_CYear = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_CYear == '#':        in_CYear = ""
        iParam=iParam+1
        in_StdPop = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_StdPop == '#':       in_StdPop = ""
        iParam=iParam+1
        in_LatBin = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_LatBin == '#':       in_LatBin = ""
        iParam=iParam+1
        in_MSNormWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_MSNormWDS == '#':    in_MSNormWDS = LDN.WDSMSNorm()
        iParam=iParam+1

        in_ScratchWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_ScratchWDS == '#':    in_ScratchWDS = LDN.WDSScratch()
        iParam=iParam+1
        in_KTO = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_KTO == '#':          in_KTO = ""
        iParam=iParam+1

        blnKTO = (in_KTO.lower() == "true")

        in_PopPeriod = "all"
        (strRPer, strRPeriod, RPerYears, RPerMonths) = LF.parsePeriod(in_ResPeriod)
        (strLPer, strLPeriod, (reqMinYrs, reqMaxYrs), LPerMonths) = LF.parsePeriod(in_LifePeriod)
        (strPPer, strPPeriod, (reqMinPage, reqMaxPage), PPerMonths) = LF.parsePeriod(in_PopPeriod)
        strCYear = LF.parseCYear(in_CYear)
        dblLatBin = float(in_LatBin)
        strLatBin = LF.parseLat(in_LatBin)

        NoLat = LDN.NoLat()

        allGs = LDN.AllGenders() # ['a', 'f', 'm']
        allUs = LDN.AllUrbans() # ['a', 'u', 'r']
        allGUs = [ g + u for g in allGs for u in allUs ]
        lutGU = dict( [ g, dict( [ u, g + u ] for u in allUs ) ] for g in allGs )
```

```

# census year groups; last group open ended
# group definitions, year limits, month limits
allCYs = LDN.AllCYYearGroups(strCYear)
allCYLFlDs = LDN.AllCYLatFieldNames(strCYear)
allCYUFlDs = LDN.AllCYUrbFieldNames(strCYear)
allCYLims = LDN.AllCYYearLimits(strCYear)
numCYs = [ allCYLims[i][1] - allCYLims[i][0] + 1 for i in range(0,len(allCYLims) - 1) ]
numCYs.append(None)
lutCYGrp = []
MaxCYGrp = len(allCYLims) - 1
# lut for years to groups - not too many years
# - too many months for equivalent month lut
for i in range(0,len(allCYLims) - 1):
    lutCYGrp.extend([i] * (numCYs[i]))
lutCYGrp.extend([len(allCYLims) - 1])
MaxCYlut = len(lutCYGrp) - 1 # should = allCYLims[len(allCYLims)-1][0]

if strPPeriod=="all":
    cyGrpLo = 0
    cyGrpHi = MaxCYGrp
else:
    # need to filter for high end open
    reqMinPage = int(reqMinPage)
    cyGrpLo = lutCYGrp[reqMinPage]
    if reqMaxPage=="":
        cyGrpHi = MaxCYGrp
    else:
        reqMaxPage = int(reqMaxPage)
        cyGrpHi = lutCYGrp[reqMaxPage]

usedPCYGrps = range(cyGrpLo,cyGrpHi+1)

# factors
# input table of factors, though only use name
if in_Factor:
    strFactorTab = in_Factor
    strFactor = gp.Describe(strFactorTab).BaseName
    allFs = LDN.AllFactors(strFactor)
else: allFs = [""]

# input table of effective latitude for each case
strELatsTabName = LDN.KeepEpochLats(strRPer, strLPer)
strELatsTab = in_EffWDS + "/" + strELatsTabName

# input table of standard population
strStdPopTab = in_StdPop
strStdPop = gp.Describe(strStdPopTab).BaseName

# intermediate table of numbers of cases per latitude bin
strBLatsTabName = {}
strBLatsTab = {}
# output table of number of MS cases normalised by census population
strNLatsTabName = {}
strNLatsTab = {}

strTmpLatTabName = {}
strTmpLatTab = {}
strTmpNormTabName = {}
strTmpNormTab = {}
for f in allFs:
    if f: sf = "_" + f
    else: sf = ""
    strBLatsTabName[f] = LDN.KeepLatENums(strRPer, strLPer, strLatBin, f)
    strBLatsTab[f] = in_EffWDS + "/" + strBLatsTabName[f]
    strNLatsTabName[f] = LDN.KeepLatEPNorm(strRPer, strLPer, strCYear, strStdPop, strLatBin, f)
    strNLatsTab[f] = in_MSNormWDS + "/" + strNLatsTabName[f]
    strTmpLatTabName[f] = "tmpLatBins" + sf
    strTmpLatTab[f] = in_ScratchWDS + "/" + strTmpLatTabName[f]
    strTmpNormTabName[f] = "tmpNorm" + sf
    strTmpNormTab[f] = in_ScratchWDS + "/" + strTmpNormTabName[f]

strTmpELView = "tmpELatNumsView"
strTmpStdView = "tmpStdPopView"

strPLatsTabName = {}
strPLatsTab = {}
strTmpPopView = {}
# input table of census population per latitude bin
for gu in allGUs:
    strPLatsTabName[gu] = LDN.KeepLatPop(strCYear, strLatBin, gu)
    strPLatsTab[gu] = in_LatPopWDS + "/" + strPLatsTabName[gu]
    strTmpPopView[gu] = "tmpPopView_" + gu

strFld_lp = {}
strFld_lno = {}
strFld_pht = {}
strFld_pse = {}
strFld_ast = {}
strFld_ase = {}
for gu in allGUs:
    strFld_lp[gu] = LDN.AllFieldsLatPop(gu, strCYear) # LDN.FieldLatPop(gu)

```

```

    strFld_lno[gu] = LDN.FieldLatNum(gu)
    strFld_pht[gu] = LDN.FieldLatNumPHT(gu)
    strFld_pse[gu] = LDN.FieldLatNumPSE(gu)
    strFld_ast[gu] = LDN.FieldLatNumAST(gu)
    strFld_ase[gu] = LDN.FieldLatNumASE(gu)

strFld_std = LDN.AllStdPopFieldNames(strCYear)
strFld_lc = LDN.FieldLatC()
strFld_ls = LDN.FieldLatS()
strFld_ln = LDN.FieldLatN()
strFld_gen = LDN.FieldGender()
strFld_eff = LDN.FieldLatEffect()
strFld_urb = LDN.FieldUrban()
strFld_fac = LDN.FieldFactor()

if not gp.Exists(strELatsTab):
    gp.AddMessage(("MS Effective Latitude Data table %s does not exist\n\n" +
                  "Please run SelectEpochData with manual input\n") % strELatsTab)
    return 0

if not gp.Exists(strPLatsTab[allGUs[0]]):
    gp.AddMessage(("Population by Latitude table %s does not exist - creating..."
                  % strPLatsTab[allGUs[0]])
    # LatitudePops_latitude
    # <PopWDS>, <CYear>, <AgeGroup>, <SliceWDS>, <LatSize>, <LatPopWDS>, <ScratchWDS>, <KTO>
    import LatitudePops
    LatitudePops.main("#", in_CYear, in_PopPeriod, "#",
                      in_LatBin, in_LatPopWDS, in_ScratchWDS, "f")
    if not gp.Exists(strPLatsTab[allGUs[0]]):
        gp.AddMessage("Population by Latitude table %s does not exist"
                      % strPLatsTab[allGUs[0]])
        gp.AddMessage("called LatitudePops with args:\n %s" % repr("#", in_CYear,
                               in_PopPeriod, "#", in_LatBin, in_LatPopWDS, in_ScratchWDS, "f"))
        return 0
    else:
        gp.AddMessage("Created.")

gp.AddMessage("Reading Epoch Data...")

# need reference lat for latitude bins
# take from latitude pops data
# sort Ascending will give the most southerly lat

get_rows = gp.SearchCursor(strPLatsTab[allGUs[0]], "", "", "", strFld_ls + " A")
get_row = get_rows.Next()

if not get_row:
    gp.AddMessage("Cannot get latitude extent from %s" % strPLatsTab[allGUs[0]])
    exit(0)

sLats = get_row.GetValue(strFld_ls)

while get_row:
    nLatN = get_row.GetValue(strFld_ln)
    get_row = get_rows.Next()

# done with the cursor
del get_row
del get_rows

numBins = int(round(((nLatN - sLats) / dblLatBin), 0))
### Don't use "*" in place of outer list comprehension ###
### - get pointer to same list rather than multiple lists! ###
intLatCounter = dict([(gu, dict([(f, [[0] * (MaxCYGrp+1) for x in range(numBins)])
                                for f in allFs])) for gu in allGUs])
dblLatC = [sLats + (float(x) + 0.5) * dblLatBin for x in range(numBins)]

# read in effective lats and bin them
# now have by census age group, not just for whole period
# reqMinYrs & reqMaxYrs are extremes of life period to be considered
# if all, prevalence or prior to onset then need all census groups
# else looking at age period (with or without pre-birth)
# and period will intersect fixed subset of census groups
#
if strLPeriod in ("onset", "prev"):
    iMinCYrGrp = 0
    iMaxCYrGrp = MaxCYGrp
else:
    iMinCYrGrp = lutCYGrp[min(reqMinYrs, MaxCYlut)]
    iMaxCYrGrp = lutCYGrp[min(reqMaxYrs, MaxCYlut)]

usedLCYGrps = range(iMinCYrGrp, iMaxCYrGrp+1)

try:
    if gp.Exists(strTmpELView):
        gp.Delete(strTmpELView)
except:
    gp.AddMessage("Error trying to check for and delete %s" % strTmpELView)
gp.MakeTableView_management(strELatsTab, strTmpELView)

# searchCursor

```

```

# create search cursor
get_rows = gp.SearchCursor(strTmpELView)
get_row = get_rows.Next()

gp.AddMessage("Binning data...")
strA = 'a'
while get_row:
    strG = get_row.GetValue(strFld_gen)
    strF = get_row.GetValue(strFld_fac)
    for nGrp in usedLCYGrps: # range(iMinCYrGrp, iMaxCYrGrp + 1):
        dblU = get_row.GetValue(allCYUFlds[nGrp])
        strU = 'u' if dblU > 0.5 else 'r'
        lat = get_row.GetValue(allCYLFLds[nGrp])

        if lat != NoLat:
            # bin number
            numBin = int( (lat - sLatS)/dblLatBin )
            if numBin < 0 or numBin >= numBins:
                fkuin = get_row.GetValue("fk_uin")
                gp.AddMessage("Got Latitude out of range!\n%s\nfk_uin = %d lat_effect = %s"
                    + " numBin = %d" % (strELatsTab, fkuin, repr(lat), numBin))
            else:
                for x in [strG, strA]:
                    for y in [strU, strA]:
                        xy = lutGU[x][y]
                        for z in [strF, strA]:
                            intLatCount[xy][z][numBin][nGrp] += 1
        get_row = get_rows.Next()

# done with the cursor
del get_row
del get_rows

# sum all age groups for crude total
intLatSum = dict([(gu, dict([(f, [sum(intLatCount[gu][f][x]) for x in range(0, numBins)])
    for f in allFs])) for gu in allGUs])

# write out binned - just save totals for lat bins
for strF in allFs:
    strTab = strTmpLatTab[strF]
    try:
        if gp.Exists(strTab): gp.Delete(strTab)
    except:
        gp.AddMessage("Error trying to check for and delete %s" % strTab)
    gp.AddMessage("Creating Temporary Table...")
    gp.CreateTable(in_ScratchWDS, strTmpLatTabName[strF])

    gp.AddMessage("Setting up new table...")
    gp.AddField(strTab, strFld_lc, "double")
    for gu in allGUs:
        gp.AddField(strTab, strFld_lno[gu], "long")

    gp.AddMessage("Getting insert cursor...")
    # create insert cursor for new table
    ins_rows = gp.InsertCursor(strTab)

    for i in range(0, numBins):
        ins_row = ins_rows.NewRow()

        ins_row.SetValue(strFld_lc, dblLatC[i])
        for gu in allGUs:
            ins_row.SetValue(strFld_lno[gu], intLatSum[gu][strF][i])

        ins_rows.InsertRow(ins_row)

    # done with the cursor
    del ins_row
    del ins_rows

    gp.CopyRows_management(strTab, strBLatsTab[strF])

# read in pop data
# now have table for each gu and field for each cy group

gp.AddMessage("Reading census population data...")

dblPopCount = dict([gu, [[0.0 for y in range(0, MaxCYGrp+1)]
    for x in range(0, numBins)] ] for gu in allGUs)

for gu in allGUs:
    try:
        if gp.Exists(strTmpPopView[gu]):
            gp.Delete(strTmpPopView[gu])
    except:
        gp.AddMessage("Error trying to check for and delete %s" % strTmpPopView[gu])
    gp.MakeTableView_management(strPLatsTab[gu], strTmpPopView[gu])

# searchCursor

get_rows = gp.SearchCursor(strTmpPopView[gu])

```

```

get_row = get_rows.Next()

while get_row:
    lat = float(get_row.GetValue(strFld_lc))
    # bin number - avoid mismatches due to rounding errors
    numBin = int( (lat - sLatS)/dblLatBin )
    for i in usedPCYGrps:
        pop = float(get_row.GetValue(strFld_lp[gu][i]))
        dblPopCount[gu][numBin][i] = pop # if pop > 0 else 1

    get_row = get_rows.Next()

# done with the cursor
del get_row
del get_rows

gp.AddMessage("Reading standard population data...")

dblStdPopCount = [0.0 for y in range(0, MaxCYGrp+1)]

gp.MakeTableView_management(strStdPopTab, strTmpStdView)

# searchCursor

get_rows = gp.SearchCursor(strTmpStdView)
get_row = get_rows.Next()

# should only be one row
for i in usedPCYGrps:
    pop = float(get_row.GetValue(strFld_std[i]))
    dblStdPopCount[i] = pop # if pop > 0 else 1

# done with the cursor
del get_row
del get_rows

# normalise data

gp.AddMessage("Normalising data...")
# need to Age Standardise... mean(ratio of each age group)
# dblNormData is ratio for each gu-bin
# dblRatioData is ratio for each gu-bin-group
# dblAgeStdData is age standardised
# ..SE's are standard errors

# crude rate
dblPopSum = dict([gu, [sum(dblPopCount[gu][x]) for x in range(0, numBins)] ]
                 for gu in allGUs)

dblNormData = dict([gu, dict([f,
    [(intLatSum[gu][f][x]/dblPopSum[gu][x] if dblPopSum[gu][x] != 0.0 else 0.0)
    for x in range(0, numBins)] ] for f in allFs)] for gu in allGUs)

# SE is sqrt(p(1-p)/n) with p=r/n being the crude rate for pop [sample] n
dblNormSE = dict([gu, dict([f,
    [(math.sqrt(dblNormData[gu][f][x] * (1.0-dblNormData[gu][f][x])
    / dblPopSum[gu][x]) if dblPopSum[gu][x] != 0.0 else 0.0)
    for x in range(0, numBins)] ] for f in allFs)] for gu in allGUs)

# Age Standardised
# - sum of std pop * age specific rates
dblAgeRatesData = dict([gu, dict([f,
    [(dblStdPopCount[y]*intLatCount[gu][f][x][y]/dblPopCount[gu][x][y]
    if intLatCount[gu][f][x][y]
    and dblPopCount[gu][x][y] != 0.0 else 0.0) for y in range(0, MaxCYGrp+1)]
    for x in range(0, numBins)] ] for f in allFs)] for gu in allGUs)

dblAStdData = dict([gu, dict([f, [sum(dblAgeRatesData[gu][f][x])
    for x in range(0, numBins)] ] for f in allFs)] for gu in allGUs)

# SE is much the same as the data but need (pop/pop)^2 then sqrt of sum
dblAgeRatesSE = dict([gu, dict([f,
    [(intLatCount[gu][f][x][y]*(dblStdPopCount[y]/dblPopCount[gu][x][y])**2
    if intLatCount[gu][f][x][y]
    and dblPopCount[gu][x][y] != 0.0 else 0.0) for y in range(0, MaxCYGrp+1)]
    for x in range(0, numBins)] ] for f in allFs)] for gu in allGUs)

dblAStdSE = dict([gu, dict([f, [math.sqrt(sum(dblAgeRatesSE[gu][f][x]))
    for x in range(0, numBins)] ] for f in allFs)] for gu in allGUs)

# create data table

for strF in allFs:
    strTab = strTmpNormTab[strF]
    if gp.Exists(strTab):
        gp.Delete(strTab)
    gp.CreateTable(in_ScratchWDS, strTmpNormTabName[strF])

    gp.AddField(strTab, strFld_ln, "double")
    gp.AddField(strTab, strFld_ls, "double")
    gp.AddField(strTab, strFld_lc, "double")

```

```

for gu in allGUs:
    gp.AddField(strTab, strFld_pht[gu], "double")
    gp.AddField(strTab, strFld_pse[gu], "double")
    gp.AddField(strTab, strFld_ast[gu], "double")
    gp.AddField(strTab, strFld_ase[gu], "double")

dblHLatBin = 0.5 * dblLatBin
dblLatN = [dblLatC[x] + dblHLatBin for x in range(0, numBins)]
dblLatS = [dblLatC[x] - dblHLatBin for x in range(0, numBins)]

# create insert cursor for new table
ins_rows = gp.InsertCursor(strTab)

dblPHT = 100000.0
for i in range(0, numBins):
    ins_row = ins_rows.NewRow()

    ins_row.SetValue(strFld_ln, dblLatN[i])
    ins_row.SetValue(strFld_ls, dblLatS[i])
    ins_row.SetValue(strFld_lc, dblLatC[i])
    for gu in allGUs:
        ins_row.SetValue(strFld_pht[gu], dblNormData[gu][strF][i] * dblPHT)
        ins_row.SetValue(strFld_pse[gu], dblNormSE[gu][strF][i] * dblPHT)
        ins_row.SetValue(strFld_ast[gu], dblAStdData[gu][strF][i])
        ins_row.SetValue(strFld_ase[gu], dblAStdSE[gu][strF][i])

    ins_rows.InsertRow(ins_row)

# done with the cursor
del ins_row
del ins_rows

# save to gdb
gp.AddMessage("Writing data...")

gp.CopyRows_management(strTab, strNLatsTab[strF])

### Done ###
gp.AddMessage("Done.")
strCodeSec = "End!"
strExtraInfo = ""

if blnKTO:
    strCodeSec = "exception to keep tool open!"
    exit() # keep tool open in case works and gets to here!

except:
    gp.AddMessage("Got exception in ProcessEpochLats, section = %s %s"
                  % (strCodeSec, strExtraInfo))
    raise
else:
    return 0 # exit errorlessly

if __name__ == '__main__':
    main(*sys.argv[1:])

```


LatitudeFunctions.py

```

# -----
# LatitudeFunctions.py
# -----
# mainly recoding and parsing functions which need to be common to all scripts
# not using an geoprocessing or tools

# defines:
#   parseCYear, parseLat, parseGender, parseFactor, parseFacType
#   parseStdPop, parsePeriod, checkPeriods

# Import system modules
import sys, string, os, arcgisscripting

# Create the Geoprocessor object
gp = arcgisscripting.create(9.3)

def parseCYear(strCYear):
    iYear = int(strCYear.strip())
    if (iYear % 5)!=1:
        gp.AddMessage("Year %d is not a census year!" % iYear)
        exit(1)
    if iYear < 1951:
        gp.AddMessage("Year %d is pre 1951!" % iYear)
        exit(1)
    if iYear > 2006:
        gp.AddMessage("Year %d is later than 2006!" % iYear)
        exit(1)
    return strCYear.strip()

def parseLat(strLatSize):
    dLat = float(strLatSize.strip())
    iLat10 = int(dLat*10.0)
    strLat = "%02d" % iLat10
    return strLat

def parseGender(strGend):
    strGen = strGend.strip() + "a" # gives "a" if blank
    strGen = strGen.strip().lower()[0]
    if strGen not in "afm": strGen = "a"
    return strGen

def parseFactor(strFactor):
# just do lower for now
    strFac = strFactor.strip().lower()
    if strFac == "#": strFac = ""
    return strFac

def parseFacType(strFacType):
# just do lower for now
    strFacT = strFacType.strip().lower()
    if strFacT == "#": strFacT = ""
    return strFacT

def parseStdPop(strStdPop):
# just do lower for now
    strStd = strStdPop.strip().lower()
    return strStd

# can't have '-' in GDB table names, so can't use e.g '0-4' as part of name
# need to re-code pop, residence and life period codes
#
#   pop res lif mnm      mxm      mny      mxy      notes
# all    all    all    all    0
# 0-4    0a4    0a4    0a4    m1    m2    y1    y2    num1 < num2    age
# 0a4    0a4    0a4    0a4    m1    m2    y1    y2    num1 < num2    age
# 5-0    -      5o0    5o0    m2    m1    y2    y1    num1 > num2    prior to onset
# 5o0    -      5o0    5o0    m2    m1    y2    y1    num1 > num2    prior to onset
# 1-1    1a1    1a1    1a1    m1    m2    y1    y2    num1 = num2    age
# 1a1    1a1    1a1    1a1    m1    m2    y1    y2    num1 = num2    age
# 1o1    -      1o1    1o1    m2    m1    y2    y1    num1 = num2    prior to onset
# c-4    -      -      ca4    -9    m2    0    y2    include pre birth
# ca4    -      -      ca4    -9    m2    0    y2    include pre birth
# c-b    -      -      cab    -9    0    0    0    only pre birth
# cab    -      -      cab    -9    0    0    0    only pre birth
# 90a    90a    -      m1    y1    high end open
# prv    -      prv    prv    prevalence
#
# function used for age period, res period and life period
# need sanitised period code before can construct table names etc
# removed where clause generation back to script

# will return limits in years and months
# e.g. 0a4 -> 0, 4 and 0, 60 - Note 'extra' 12 months in max

def parsePeriod(strPerd):
#   fac = 12 if strYorM.strip().upper()[0] == "M" else 1
    strPer = strPerd.strip().lower()

```

```

if strPer == "": strPer = "all"
if "-" in strPer:
    # change '-' to 'a' unless have decreasing range => onset 'o'
    (reqMinYrs, reqMaxYrs) = strPer.split("-")
    strPer = reqMinYrs + "a" + reqMaxYrs
    if reqMinYrs.isdigit() and reqMaxYrs.isdigit():
        if reqMinYrs > reqMaxYrs: strPer = reqMinYrs + "o" + reqMaxYrs
# should now have only valid recoded codes
strCode = strPer
if strPer=="all":
    posn = 0 # all epochs from birth to onset
    strPeriod = "all"
    reqMinMonth = 0
    reqMaxMonth = ""
    reqMinYear = 0
    reqMaxYear = ""
elif strPer=="prv":
    posn = 7 # special prevalence point for life
    strPeriod = "prev"
    reqMinMonth = ""
    reqMaxMonth = ""
    reqMinYear = ""
    reqMaxYear = ""
elif strPer[0] == "c":
    posn = 1 # have period from conception
    reqMinMonth = -9
    reqMinYear = 0
    if strPer[2] == "b":
        posn = 3 # have just pre-birth period
        strPeriod = "preb"
        reqMaxMonth = 0
        reqMaxYear = 0
    else:
        posn = 4 # have pre-birth plus age period
        strPeriod = "pbage"
        reqMaxYear = int(strPer[2:])
        reqMaxMonth = 12 * (reqMaxYear + 1)
else:
    # have age period or period prior to onset
    posn = 2
    if "o" in strPer:
        # have period prior to onset
        posn = 5
        strPeriod = "onset"
        (reqMaxYear, reqMinYear) = strPer.split("o")
    else:
        # have age period
        posn = 6
        strPeriod = "age"
        (reqMinYear, reqMaxYear) = strPer.split("a")
        reqMinYear = int(reqMinYear)
        reqMinMonth = 12 * int(reqMinYear)
# might have open ended for population
if reqMaxYear=="":
    reqMaxMonth = ""
else:
    reqMaxYear = int(reqMaxYear)
    reqMaxMonth = 12 * (reqMaxYear + 1)
return (strCode, strPeriod, (reqMinYear, reqMaxYear), (reqMinMonth, reqMaxMonth))

# need to have lper as subset of res (unless prev)
# probably only used in epoch script but here for completeness
#
# allowed combinations
#
# lper    prv all age    onset    incl-pb    only-pb
# res
# all      OK      OK      OK      OK      OK      OK
# age      OK      NO      SS      NO      SSB     B
# onset    OK      NO      NO      NO      SS      NO      NO
# prv      OK     NO     NO     NO     NO     NO
#
# OK = combination allowed - no further check
# NO = combination not allowed - no further check
# SS = combination allowed if lper age is subset of res
# B = combination allowed if res age from 0
# SSB = SS+B

# function used to check res & lper combination allowed
# period codes already sanitised

def checkPeriods(strRPer, strRPerD, strLPer, strLPerD):
    if strRPerD == "all": isOK = True
    elif strLPerD == "prev": isOK = True
    elif strLPerD == "all": isOK = False
    elif strRPerD == "prev": isOK = False
    elif strRPerD == "onset":
        if strLPerD != "onset": isOK = False
        else:
            # onset but need to check limits
            (rHi,rLo) = strRPer.split("o")
            (lHi,lLo) = strLPer.split("o")
            if int(lLo) < int(rLo) or int(lHi) > int(rHi): isOK = False
            else: isOK = True
    else:
        # have res = age
        if strLPerD == "onset": isOK = False

```

```
else:          # age but need to check limits
    (rLo,rHi) = strRPer.split("a")
    (lLo,lHi) = strLPer.split("a")
    if lLo == "c": lLo = "0"
    if lHi == "b": lHi = "0"
    if int(lLo) < int(rLo) or int(lHi) > int(rHi): isOK = False
    else: isOK = True

return isOK
```

LatitudeDataNames.py

```
# -----
# LatitudeDataNames.py
# -----
# encoding of data names which need to be common to various scripts
# not using any geoprocessing tools

# defines:
#
# Utility
#
#     NoLat, Max_65p, Max_85p, Max_AgeGrp
#
# Lists
#
#     AllGenders, AllUrbans
#     AllFactorNames, AllFactors, AllGUFacLabels, AllFactorLabels
#     AllYearGroups, AllYearPeriods, AllYearLimits, AllCYMthLimits
#
# Feature Class and Table names
#
#     KeepEpochCulms, KeepEpochPrevs, KeepEpochs, KeepEpochLats, KeepLatENums
#     KeepEpochHist, KeepEpochStats, KeepEpochSunHrs, KeepEpochUVdat, KeepEpochMUVdat
#     KeepPop, KeepLatPop, KeepLatGrid, KeepStatAreas, KeepStatUrban, KeepSlicedAreas
#     KeepLatEPNorm, KeepStats
#     SplitLatENums, SplitLatEPNorm
#
#     KeepGraph, KeepGraphPdf, KeepGraphNoExt
#     KeepGraphLeg, KeepGraphSumm, KeepGraphLat, KeepGraphMap
#     KeepStatsX, KeepCrossStatsX, KeepGraphLegX, KeepGraphMapX, KeepGraphPdfX
#
# Field Names
#
#     AllCYGFieldNames, AllCYGFldNames
#     AllCYLatFieldNames, AllCYUrbFieldNames, AllCYPopFieldNames, AllStdPopFieldNames
#     FieldGender, FieldLatEffect
#     FieldLatNum, FieldLatNumPHT, FieldLatNumAST, FieldLatNumPSE, FieldLatNumASE
#     FieldAreaFrac, AllFieldsSamplePop, AllFieldsLatPop, FieldsPopArea, FieldStatArea
#     FieldLatN, FieldLatS, FieldLatC, FieldListLats
#     FieldSqKm, FieldUrban, FieldYear, FieldNumYears, FieldOnsetAge
#     FieldFactor, FieldFacUin, FieldCYear, FieldStdPop, FieldResPer, FieldLifePer
#     FieldStatsCount, FieldStatsSum, FieldStatsMin, FieldStatsMax
#     FieldStatsMean, FieldStatsStdDev, FieldStatsKurt, FieldStatsSkew
#     FieldSunHours, FieldLatSunHrs, FieldSolExp
#     FieldCSUV, FieldEstVD, FieldEstUVB, FieldCSUvm, FieldEstVdm, FieldEstUVBm
#
# Workspaces or Datasets
#
#     WDSBase, WDSGrid, WDSPop, WDSAreas, WDSSlice
#     WDSLatPop, WDSStdPop, WDSMSLat, WDSMSNorm, WDSMSSun, WDSMSUV
#     WDSStats, WDSGraph, DSScratch
#
# Some field and file names now contain Gender or Gender-Urban specification
# changed arg name etc - doesn't affect defs, just clarifies

# Import system modules
import sys, string, os, arcgisscripting

# Create the Geoprocessor object
gp = arcgisscripting.create(9.3)

### utility functions

def NoLat(): return 999

### 'constant' data

def AllGenders(): return ['a', 'f', 'm']

def AllUrbans(): return ['a', 'u', 'r']

def AllFactorNames():
    return ['no_factors', 'birth_season_a', 'birth_season_b', 'birth_month',
            'ms_type', 'ms_type23', 'ms_family', 'ms_fam_p', 'ms_fam_m', 'ms_fam_f',
            'ms_fam_s', 'ms_fam_if', 'ms_fam_im', 'sd_ethnicity',
            'bo_migration', 'bo_migrated2', 'bo_migrated1',
            'bp_migration', 'bp_migrated2', 'bp_migrated1'
            ]

def AllFactors(strFacName):
# 'a' reserved for 'all' - should be included for each
    dic = {
        'no_factors': ['a', '', 'x'],
        'birth_season_a': ['a', 'w', 's', 'x'],
        'birth_season_b': ['a', 'w', 's', 'x'],
        'birth_month': ['a', '1', '2', '3', '4', '5', '6', '7', '8', '9', 'j', 'k', 'l'],
        'ms_type': ['a', 's', 'p', 'r', 'x'],
    }
```

```

'ms_type23': ['a', 'm', 'p', 'x'],
'ms_type_birthing_season': ['a', 's', 'c', 'w', 'v', 'f', 'r', 'x'],
'ms_family': ['a', 'i', 'd', 'x'],
'ms_fam_p': ['a', 'b', 'm', 'f', 'x'],
'ms_fam_m': ['a', 'm', 'x'],
'ms_fam_f': ['a', 'f', 'x'],
'ms_fam_s': ['a', 'f', 'h', 'x'],
'ms_fam_if': ['a', 'f', 'x'],
'ms_fam_im': ['a', 'm', 'x'],
'sd_ethnicity': ['a', 'e', 'm', 'x'],
'bo_migration': ['a', 'n', 's', 'x'],
'bo_migrated2': ['a', 'm', 'x'],
'bo_migrated1': ['a', 'm', 'x'],
'bp_migration': ['a', 'n', 's', 'x'],
'bp_migrated2': ['a', 'm', 'x'],
'bp_migrated1': ['a', 'm', 'x']
}
if dic.has_key(strFacName):
    return dic[strFacName]
else:
    return ['']

def AllGUFacLabels():
    return {'aa': "All Pop", 'au': "All Urban Pop", 'ar': "All Rural Pop",
            'fa': "Female Pop", 'fu': "Female Urban Pop", 'fr': "Female Rural Pop",
            'ma': "Male Pop", 'mu': "Male Urban Pop", 'mr': "Male Rural Pop"}

def AllFactorLabels(strFacName):
    dic = {
        'no_factors': { 'a': "All cases",
                        '': "All no_factor cases",
                        'x': "No cases"},
        'birth_season_a': { 'a': "All cases",
                            's': "Born Summer (Dec-Jan)",
                            'w': "Born Winter (Jun-Jul)",
                            'x': "Born Spring/Autumn (Feb-May, Aug-Nov)"},
        'birth_season_b': { 'a': "All cases",
                            's': "Born Summer (Nov-Feb)",
                            'w': "Born Winter (May-Aug)",
                            'x': "Born Spring/Autumn (Mar-Apr, Sep-Oct)"},
        'birth_month': { 'a': "All cases",
                          '1': "Born January",
                          '2': "Born February",
                          '3': "Born March",
                          '4': "Born April",
                          '5': "Born May",
                          '6': "Born June",
                          '7': "Born July",
                          '8': "Born August",
                          '9': "Born September",
                          'j': "Born October",
                          'k': "Born November",
                          'l': "Born December"},
        'ms_type': { 'a': "All MS Phenotypes",
                     's': "SPMS MS Phenotype",
                     'p': "PPMS MS Phenotype",
                     'r': "RRMS MS Phenotype",
                     'x': "Other or Unspecified MS Phenotype"},
        'ms_type23': { 'a': "All MS Phenotypes",
                       'm': "RR/SPMS MS Phenotypes",
                       'p': "PPMS MS Phenotype",
                       'x': "Other or Unspecified MS Phenotype"},
        'ms_type_birthing_season': { 'a': "All cases",
                                      's': "RR/SPMS Born Summer (Nov-Feb)",
                                      'c': "PPMS Born Summer (Nov-Feb)",
                                      'w': "RR/SPMS Born Winter (May-Aug)",
                                      'v': "PPMS Born Winter (May-Aug)",
                                      'f': "RR/SPMS Born Spring/Autumn (Mar-Apr, Sep-Oct)",
                                      'r': "PPMS Born Born Spring/Autumn (Mar-Apr, Sep-Oct)",
                                      'x': "Other or Unspecified combination"},
        'ms_family': { 'a': "All cases",
                       'i': "MS in immediate family",
                       'd': "MS in distant but not immediate family",
                       'x': "No MS in rest of family"},
        'ms_fam_p': { 'a': "All cases",
                      'b': "Both parents have MS",
                      'm': "Mother has MS, Father does not",
                      'f': "Father has MS, Mother does not",
                      'x': "Neither parents have MS"},
        'ms_fam_m': { 'a': "All cases",
                      'm': "Mother has MS",
                      'x': "Mother does not have MS"},
        'ms_fam_f': { 'a': "All cases",
                      'f': "Father has MS",
                      'x': "Father does not have MS"},
        'ms_fam_s': { 'a': "All cases",
                      'f': "At least one full sibling has MS",
                      'h': "At least one half sibling has MS",
                      'x': "No full or half siblings have MS"},
        'ms_fam_if': { 'a': "All cases",
                       'f': "MS in female immediate family members",

```

```

        'x': "No MS in female immediate family members"},
    'ms_fam_im': {
        'a': "All cases",
        'm': "MS in male immediate family members",
        'x': "No MS in male immediate family members"},
    'sd_ethnicity': {
        'a': "All Ethnicities",
        'e': "NZ Euro or Euro Ethnicity",
        'm': "Maori or PI Ethnicity",
        'x': "Other or Unspecified Ethnicity"},
    'bo_migration': {
        'a': "All cases",
        's': "Moved South more than 2 deg prior to onset",
        'n': "Moved North more than 2 deg prior to onset",
        'x': "Moved less than 2 deg"},
    'bo_migrated1': {
        'a': "All cases",
        'm': "Moved more than 1 deg prior to onset",
        'x': "Moved less than 1 deg prior to onset"},
    'bo_migrated2': {
        'a': "All cases",
        'm': "Moved more than 2 deg prior to onset",
        'x': "Moved less than 2 deg prior to onset"},
    'bp_migration': {
        'a': "All cases",
        's': "Moved South more than 2 deg prior to prev",
        'n': "Moved North more than 2 deg prior to prev",
        'x': "Moved less than 2 deg"},
    'bp_migrated1': {
        'a': "All cases",
        'm': "Moved more than 1 deg prior to prev",
        'x': "Moved less than 1 deg prior to prev"},
    'bp_migrated2': {
        'a': "All cases",
        'm': "Moved more than 2 deg prior to prev",
        'x': "Moved less than 2 deg prior to prev"}
    }
    if dic.has_key(strFacName):
        return dic[strFacName]
    else:
        return {'': ""}

# different years have different max age groups
def Max_65p(): return "65+"

def Max_85p(): return "85+"

def Max_AgeGrp(strYear):
    if strYear=="2006":
        max = Max_85p()
    else:
        max = Max_65p()
    return max

def AllCYearGroups(strCYr):
    grps = [ "0-4", "5-9", "10-14", "15-19", "20-24",
             "25-29", "30-34", "35-39", "40-44", "45-49",
             "50-54", "55-59", "60-64"]
    if Max_AgeGrp(strCYr)==Max_85p():
        grps.extend(["65-69", "70-74", "75-79", "80-84", "85-"])
    else:
        grps.extend(["65-"])
    return grps

def AllCYearPeriods(strCYr):
    pers = [ "0a4", "5a9", "10a14", "15a19", "20a24",
             "25a29", "30a34", "35a39", "40a44", "45a49",
             "50a54", "55a59", "60a64"]
    if Max_AgeGrp(strCYr)==Max_85p():
        pers.extend(["65a69", "70a74", "75a79", "80a84", "85a"])
    else:
        pers.extend(["65a"])
    return pers

def AllCYearLimits(strCYr):
    # limits in terms of inclusive year numbers
    lims = [ (0,4), (5,9), (10,14), (15,19), (20,24),
             (25,29), (30,34), (35,39), (40,44), (45,49),
             (50,54), (55,59), (60,64)]
    if Max_AgeGrp(strCYr)==Max_85p():
        lims.extend([(65,69), (70,74), (75,79), (80,84), (85,None)])
    else:
        lims.extend([(65,None)])
    return lims

def AllCYMthLimits(strCYr):
    # limits in terms of month limits
    mons = [ (0,60), (60,120), (120,180), (180,240), (240,300),
             (300,360), (360,420), (420,480), (480,540), (540,600),
             (600,660), (660,720), (720,780)]
    if Max_AgeGrp(strCYr)==Max_85p():
        mons.extend([(780,840), (840,900), (900,960), (960,1020), (1020,None)])
    else:
        mons.extend([(780,None)])
    return mons

# define this here rather than below with other fields
def AllCYGFieldNames(strXYZ, strCYr):
    strGrps = AllCYearPeriods(strCYr)

```

```

    strNames = map(lambda x: "%s_%s" % (strXYZ,x), strGrps)
    return strNames

def AllCYGFldNames(strXYZ,strCYr):
# slightly shorter - without the '_' separator
    strGrps = AllCYearPeriods(strCYr)
    strNames = map(lambda x: "%s%s" % (strXYZ,x), strGrps)
    return strNames

def AllCYLatFieldNames(strCYr):
    return AllCYGFldNames("lat",strCYr)

def AllCYUrbFieldNames(strCYr):
    return AllCYGFldNames("urb",strCYr)

def AllCYPopFieldNames(strCYr):
    return AllCYGFldNames("pop",strCYr)

def AllStdPopFieldNames(strCYr):
    return AllCYGFldNames("std",strCYr)

### Feature Class and Table names

# parameters are all pre-coded

### data derived from MS epochs (e_)

### used by SelectEpochData.py
def KeepEpochCulms():
# table of all epochs fixed for onset and culms calculated
    strName = "e_culms"
    return strName

### used by SelectEpochData.py
def KeepEpochPrevs():
# table of all prevalence epochs
    strName = "e_prevs"
    return strName

### used by SelectEpochData.py
def KeepEpochs(strResPeriod):
# feature class of epochs by residence criteria
    strName = "e_%s_res" % (strResPeriod)
    return strName

### used by SelectEpochData.py
def KeepEpochHist(strResPeriod, strLifePeriod):
# table (histogram) of case years by residence criteria and life period
    strName = "e_%s_%s_hist" % (strResPeriod, strLifePeriod)
    return strName

### used by SelectEpochData.py
def KeepEpochStats(strResPeriod, strLifePeriod):
# table of summary statistics of histogram of case years
    strName = "e_%s_%s_stats" % (strResPeriod, strLifePeriod)
    return strName

### used by SelectEpochData.py, ProcessEpochLats.py
def KeepEpochLats(strResPeriod, strLifePeriod):
# table of effective latitude per case by residence criteria and life period
    strName = "e_%s_%s_lats" % (strResPeriod, strLifePeriod)
    return strName

### used by ProcessEpochLats.py, PlotHists.py
def KeepLatENums(strResPeriod, strLifePeriod, strLatBin, strFactor):
# table of number of cases per latitude bin by residence criteria, life period and factor
    if strFactor: strF = "_" + strFactor
    else: strF = ""
    strName = "e_%s_%s_%s_hist" % (strResPeriod, strLifePeriod, strLatBin, strF)
    return strName

### sun data derived from a combination of sources (n_)
### used by

def KeepEpochSunHrs(strResPeriod, strLifePeriod, nWMth, nSMth):
# table of sun hours per case by residence criteria and life period
    strName = "e_%s_%s_%d_sun" % (strResPeriod, strLifePeriod, nWMth, nSMth)
    return strName

def KeepEpochUVdat(strResPeriod, strLifePeriod, nWMth, nSMth):
# table of UV hours per case by residence criteria and life period
    strName = "e_%s_%s_%d_uv" % (strResPeriod, strLifePeriod, nWMth, nSMth)
    return strName

def KeepEpochMUVdat(strResPeriod, strLifePeriod, nWMth, nSMth):
# table of UV hours per case by residence criteria and life period
    strName = "e_%s_%s_%d_muv" % (strResPeriod, strLifePeriod, nWMth, nSMth)
    return strName

### data derived from census statistical area populations (c_)

```

```

### used by SelectPopData.py, LatitudePops.py
def KeepPop(strCYear):
# population data by census year and age group
    strName = "c_%s_pop" % (strCYear)
    return strName

### used by LatitudePops.py, ProcessEpochLats.py, PlotHists.py
def KeepLatPop(strCYear, strLatBin, strGU):
# population data by census year and age group, re-aggregated by latitude
    strName = "c_%s_%s_%s_pop" % (strCYear, strLatBin, strGU)
    return strName

### feature classes of polygon areas (a_)

### used by SelectEpochData.py, SliceAreas.py, CreateLatitudeGrid.py
def KeepLatGrid(strLatBin):
# feature class of latitude grid, or full extent for '00'
    strName = "a_%s_grid" % (strLatBin)
    return strName

### used by SelectAreas.py, SliceAreas.py
def KeepStatAreas(strCYear):
# feature class of census statistical areas for given year
    strName = "a_%s_areas" % (strCYear)
    return strName

### used by SliceAreas.py, LatitudePops.py
def KeepSlicedAreas(strCYear, strLatBin):
# feature class of census statistical areas for given year sliced by latitude
    strName = "a_%s_%s_slices" % (strCYear, strLatBin)
    return strName

### used by SelectAreas.py, SelectEpochData.py
def KeepStatUrban(strCYear):
# feature class of urban census statistical areas for given year
# dissolved to give one multipart feature
    strName = "a_%s_urban" % (strCYear)
    return strName

### normalised data derived from a combination of sources (n_)
### used by ProcessEpochLats.py, PlotHists.py

def KeepLatEPNorm(strResPeriod, strLifePeriod, strCYear, strStdPop, strLatBin, strFactor):
# table of number of MS cases per latitude bin normalised by population data
# need to keep the length shorter
# else ArcMap & Catalog can't see the tables
# even though they exist if use Access!
#
# by case residence criteria and life period, census year and age group
    if strFactor: strF = "_" + strFactor
    else: strF = ""
    strName = "n%s%s_%s_%s%s%s"%(strCYear,strStdPop,strLatBin,strResPeriod,strLifePeriod,strF)
    return strName

def SplitLatEParams(strName):
# given table name, split out orig params, assumes:
# strName = "e_%s_%s_%s_hist" % (strResPeriod, strLifePeriod, strLatBin, strF)

    strResPeriod = strName[2:5] # assumes 3 char res period
    if strName[-7] == "_":
        strLatBin = strName[-9:-7] # assumes 2 char lat bin
        strLifePeriod = strName[6:-10] # length can vary
        strFactor = strName[-6]
    else:
        strLatBin = strName[-7:-5] # assumes 2 char lat bin
        strLifePeriod = strName[6:-8] # length can vary
        strFactor = ""
    return (strResPeriod, strLifePeriod, strLatBin, strFactor)

def SplitLatEPNorm(strName):
# given table name, split out orig params, assumes:
# strName = "n%s%s_%s_%s%s%s"%(strCYear,strStdPop,strLatBin,strResPeriod,strLifePeriod,strF)

    strCYear = strName[1:5] # assumes 4 digit year
    strStdPop = strName[5:9] # assumes 4 char std pop
    strLatBin = strName[10:12] # assumes 2 char lat bin
    strResPeriod = strName[13:16] # assumes 3 char res period
    if strName[-2] == "_":
        strLifePeriod = strName[16:-2] # length can vary
        strFactor = strName[-1]
    else:
        strLifePeriod = strName[16:] # length can vary
        strFactor = ""
    return (strResPeriod, strLifePeriod, strCYear, strStdPop, strLatBin, strFactor)

### statistical data derived from a combination of sources (s_)
### used by PlotHists.py

### still to determine outputs for stats analysis

def KeepStats(strResPeriod, strLifePeriod, strCYear, strStdPop):

```



```

# table of stats of fits to normalised data
# need to keep the length shorter
# else ArcMap & Catalog can't see the tables
# even though they exist if use Access!
#
# by case residence criteria and life period, census year and age group
strName = "f%s%s_%s%s_stats" % (strCYear, strStdPop, strResPeriod, strLifePeriod)
return strName

### graphical data derived from a combination of sources (g_)
### used by PlotHists.py
### anticipate:

def KeepGraphLeg(strGU, strResPeriod, strLifePeriod, strCYear, strStdPop):
# legend for page of graphs of crude incidence and age standardised MS cases with fits
# by gender, case residence criteria and life period, census year and age group
strName = "g%s%s_%s_%s%s_leg.png" % (strCYear, strStdPop, strGU, strResPeriod, strLifePeriod)
return strName

def KeepGraphSumm(strGU, strResPeriod, strLifePeriod, strCYear, strStdPop):
# graph of ...
# by gender, case residence criteria and life period, census year and age group
strName = "g%s%s_%s_%s%s_summ.png" % (strCYear, strStdPop, strGU, strResPeriod, strLifePeriod)
return strName

def KeepGraphLat(strGU, strResPeriod, strLifePeriod, strCYear, strStdPop, strLatBin):
# graph of crude incidence and age standardised MS cases with fits
strName = "g%s%s_%s_%s_%s%s_fit.png" % (strCYear, strStdPop,
strLatBin, strGU, strResPeriod, strLifePeriod)
return strName

def KeepGraphMap(strGU, strCYear, strStdPop):
strName = "m%s%s_%s.png" % (strCYear, strStdPop, strGU)
return strName

def KeepGraph(strGU, strResPeriod, strLifePeriod, strCYear, strStdPop):
strName = "g%s%s_%s_%s%s_mont.png"%(strCYear,strStdPop,strGU,strResPeriod,strLifePeriod)
return strName

def KeepGraphPdf(strGU, strResPeriod, strLifePeriod, strCYear, strStdPop):
# pdf of graphs of normalised MS cases with tables of fits and summary statistics
strName = "p%s%s_%s_%s%s.pdf" % (strCYear, strStdPop, strGU, strResPeriod, strLifePeriod)
return strName

def KeepGraphNoExt(strGU, strFac, strResPeriod, strLifePeriod, strCYear, strStdPop, strLatBin):
# graph of crude incidence and age standardised MS cases with fits
# ### without file extension ###
if strFac: strF = "_" + strFac
else: strF = ""
strName = "g%s%s_%s_%s%s_%s%s_fit" % (strCYear, strStdPop, strLatBin, strGU, strF,
strResPeriod, strLifePeriod)
return strName

# timestamp versions - X -> multiple params

def KeepStatsX(strStamp): strName = "f%s_stats" % (strStamp); return strName

def KeepCrossStatsX(strStamp): strName = "x%s_stats" % (strStamp); return strName

def KeepGraphLegX(strStamp): strName = "g%s_leg.png" % (strStamp); return strName

def KeepGraphMapX(strStamp): strName = "g%s_map.png" % (strStamp); return strName

def KeepGraphPdfX(strStamp): strName = "g%s.pdf" % (strStamp); return strName

### Field Names

# gender - use explicit m/f instead of boolean - clearer!
def FieldGender(): strName = "gender"; return strName

# effective latitude
def FieldLatEffect(): strName = "lat_effect"; return strName

# number at latitude
def FieldLatNum(strGU): strName = "%s_lat_num" % strGU; return strName

# number at latitude per 100,000
def FieldLatNumPHT(strGU): strName = "%s_lat_num_pht" % strGU; return strName

# SE of number at latitude per 100,000
def FieldLatNumPSE(strGU): strName = "%s_pht_se" % strGU; return strName

# number at latitude age standardised per 100,000
def FieldLatNumAST(strGU): strName = "%s_lat_num_ast" % strGU; return strName

# SE of number at latitude age standardised per 100,000
def FieldLatNumASE(strGU): strName = "%s_ast_se" % strGU; return strName

# ratio of areas of sliced statistical area to full statistical area
def FieldAreaFrac(): strName = "area_frac"; return strName

```

```

def AllFieldsSamplePop(strXYZ,strCYr):
# population extracted from census population data
strGrps = AllCYPopFieldNames(strCYr)
strNames = map(lambda x: "%s_%s" % (strXYZ,x), strGrps)
return strNames

def AllFieldsLatPop(strXYZ,strCYr):
# population extracted from census population data
strGrps = AllCYLatFieldNames(strCYr)
strNames = map(lambda x: "%s_%s" % (strXYZ,x), strGrps)
return strNames

# name of Statistical Area field in population data
# represents the same as FieldStatArea but will join so set different
def FieldSPopArea(): strName = "c_stat_area"; return strName

# name of Statistical Area field in feature class
# represents the same as FieldStatArea but will join so set different
def FieldStatArea(): strName = "stat_area"; return strName

# northern limit of the latitude slice
def FieldLatN(): strName = "Lat_N"; return strName

# southern limit of the latitude slice
def FieldLatS(): strName = "Lat_S"; return strName

# centre of the latitude slice
def FieldLatC(): strName = "Lat_C"; return strName

def FieldListLats():
# the Lat fields strung together
strName = "%s;%s;%s" % (FieldLatN(), FieldLatS(), FieldLatC())
return strName

# physical area of statistical areas
def FieldSqKm(): strName = "sq_km"; return strName

# Urban/Rural designation of statistical areas or MS case years
def FieldUrban(): strName = "urban"; return strName

# Factor field in factor table and saved stats
def FieldFactor(): strName = "factor"; return strName

# fk_uin field in factor table
def FieldFacUin(): strName = "fk_uin_fac"; return strName

# census year in saved stats
def FieldCYear(): strName = "c_year"; return strName

# census year in saved stats
def FieldStdPop(): strName = "std_pop"; return strName

# res period in saved stats
def FieldResPer(): strName = "res_period"; return strName

# life period in saved stats
def FieldLifePer(): strName = "life_period"; return strName

# year in epoch years histogram
def FieldYear(): strName = "epoch_year"; return strName

# count of years in histogram - keep short - will get stat functions added
def FieldNumYears(strFac): strName = "num_%s_yrs" % strFac; return strName

# Age at Onset in years
def FieldOnsetAge(): strName = "onset_age"; return strName

def FieldStatsCount(fld): strName = "count_" + fld; return strName
def FieldStatsSum(fld): strName = "sum_" + fld; return strName
def FieldStatsMin(fld): strName = "min_" + fld; return strName
def FieldStatsMax(fld): strName = "max_" + fld; return strName
def FieldStatsMean(fld): strName = "mean_" + fld; return strName
def FieldStatsStdDev(fld): strName = "stddev_" + fld; return strName
def FieldStatsKurt(fld): strName = "kurt_" + fld; return strName
def FieldStatsSkew(fld): strName = "skew_" + fld; return strName

def FieldSunHours(season): strName = season + "_sunhrs"; return strName
def FieldLatSunHrs(season): strName = season + "_lathrs"; return strName
def FieldSolExp(season): strName = season + "_solexp"; return strName
def FieldCSUV(season): strName = season + "_csuv"; return strName

```

```

def FieldEstVD(season):      strName = season + "_estvd"; return strName
def FieldEstUVB(season):    strName = season + "_estuvb"; return strName
def FieldCSUVm(month):     strName = "m%02d_csuv" % month; return strName
def FieldEstVDM(month):    strName = "m%02d_estvd" % month; return strName
def FieldEstUVBm(month):   strName = "m%02d_estuvb" % month; return strName

### Default Workspace or Dataset

# should probably use environment variables or the like but this will do for now
# note path separator is '/' or '\\'

# base folder or directory for the others - including trailing separator
def WDSBase(): strName = "D:/GIS/"; return strName

# latitude grids
def WDSGrid(): strName = WDSBase() + "LatGrid.mdb"; return strName

# census populations
def WDSPop(): strName = WDSBase() + "CensusPops.mdb"; return strName

# census statistical areas
def WDSAreas(): strName = WDSBase() + "CensusAreas.mdb"; return strName

# sliced statistical areas
def WDSSTlice(): strName = WDSBase() + "SlicedAreas.mdb"; return strName

# latitude populations
def WDSLTPop(): strName = WDSBase() + "LatPops.mdb"; return strName

# latitude populations
def WDSSTDPop(): strName = WDSBase() + "SlicedPops.mdb"; return strName

# MS effective latitude data
def WDSMSLat(): strName = WDSBase() + "EpochData.mdb"; return strName

# normalised MS data
def WDSMSNorm(): strName = WDSBase() + "NormData.mdb"; return strName

# MS sun hours data
def WDSMSSun(): strName = WDSBase() + "SunData.mdb"; return strName

# MS UV hours data
def WDSMSUV(): strName = WDSBase() + "UVData.mdb"; return strName

# results of statistical analyses of normalised MS data
def WDSStats(): strName = WDSBase() + "Fits.mdb"; return strName

# graphs of normalised MS data and summary analyses
def WDSGraph(): strName = WDSBase() + "Png"; return strName

# scratch
def WDSScratch(): strName = "D:/Workspace/Scratch.mdb"; return strName

```

BatchFactors.py

```
# -----
# BatchFactors.py
# -----

# script doesn't use parameters
# - set manually for data required

import arcgisscripting, ProcessEpochLats

# Create the Geoprocessor object
gp = arcgisscripting.create(9.3)

import SelectEpochData, ProcessEpochLats, PlotGraphsX4
import LatitudeDataNames as LDN

res_prfs = ["res_all"]
# if only doing select just use one std - not used but in loop
std_pops = [ "euro" ]

lat_bins = ["4.0", "3.0", "2.0"]

ms_factors = [ "no_factors" ]
# must be in list below...

ms_fac_grs = { "no_factors": (1, "a", "No Factor"),
               "birth_season_a": (4, "a;i;x;s", "Comparing (narrow) Season of Birth"),
               "birth_season_b": (4, "a;w;x;s", "Comparing (wide) Season of Birth"),
               "birth_month": (4, "1;2;3;4", "Comparing Month of Birth"),
               "sd_ethnicity": (4, "a;e;x;m", "Comparing self-declared ethnicity"),
               "ms_type": (4, "a;p;x;s", "Comparing MS phenotypes"),
               "ms_type23": (3, "a;m;p", "Comparing MS phenotypes"),
               "ms_type_birth_season": (4, "s;c;w;i;v",
                                       "Comparing MS phenotypes and (wide) Season of Birth"),
               "ms_family": (4, "a;i;x;d", "Comparing MS in family"),
               "ms_fam_p": (4, "x;m;b;f", "Comparing MS in parents"),
               "ms_fam_m": (3, "a;m;x", "Comparing mother with MS"),
               "ms_fam_f": (3, "a;f;x", "Comparing father with MS"),
               "ms_fam_s": (4, "a;f;x;h", "Comparing MS amongst siblings"),
               "ms_fam_if": (3, "a;f;x", "Comparing females with MS in immediate family"),
               "ms_fam_im": (3, "a;m;x", "Comparing males with MS in immediate family"),
               "bo_migration": (4, "a;n;x;s", "Comparing lifetime migration to onset"),
               "bo_migrated1": (3, "a;m;x", "Comparing lifetime migrated to onset"),
               "bo_migrated2": (3, "a;m;x", "Comparing lifetime migrated to onset"),
               "bp_migration": (4, "a;n;x;s", "Comparing lifetime migration to prevalence"),
               "bp_migrated1": (3, "a;m;x", "Comparing lifetime migrated to prevalence"),
               "bp_migrated2": (3, "a;m;x", "Comparing lifetime migrated to prevalence")
               }

# "birth_month" dummy entry for process - have 12 months
# for plot change to four factors required:
#       "birth_month": (4, "1;2;3;4", "Comparing Month of Birth"),
#       "birth_month": (4, "5;6;7;8", "Comparing Month of Birth"),
#       "birth_month": (4, "9;j;k;l", "Comparing Month of Birth"),

gu_fac = [ "aa", "fa", "ma" ]

gu_lab_grs = { 'aa': "All", 'fa': "Females", 'ma': "Males" }

# res & life periods & year combos:
pers_yrs = [ ("all", "0a4", "1956"), ("all", "prv", "2006") ]

# which processes to do
blnSelect = True # False
blnProcess = True # False
blnGraphs = False # True

strBaseDir = "D:/GIS/Data"

# in_memory 7-10 times faster than to mdb !!!
strScratchWDS = "in_memory"
# also need non-in_memory for select since already uses in_memory
strSelScratchWDS = "in_memory;D:/Workspace/Scratch.mdb"

# if need to save & look at temp tables then switch in_memory to scratch
#strScratchWDS = "D:/Workspace/Scratch.mdb"
#strSelScratchWDS = "D:/Workspace/Scratch.mdb;D:/Workspace/Scratch.mdb"

strResDB = "/".join([strBaseDir,"MS_res_data.mdb"])
strGridWDS = "/".join([strBaseDir,"LatGrid.mdb", "NZ_deg"])
strUrbWDS = "/".join([strBaseDir,"CensusAreas.mdb", "NZ_map"])

strPopLats = "/".join([strBaseDir,"LatPops.mdb"])

# tag output if need e.g. "_x"
strOutTag = ""
strGrDir = "graphs%s" % strOutTag

for strPref in res_prfs:
```

```

strPrefWDS = "/" .join([strBaseDir,"tmp_PF.mdb", strPref])
for strStd in std_pops:
    strStdPopWDS = "/" .join([strBaseDir,"tmp_SP.mdb", strStd])

    for strFac in ms_factors:
        strEpochWDS = "/" .join([strBaseDir,strPref,strFac,"tmp_ED.mdb"])
        strNormWDS = "/" .join([strBaseDir,strPref,strFac,"tmp_ND.mdb"])
        strFacsWDS = (strFac and "/" .join([strBaseDir,"tmp_FF.mdb", strFac]) ) or "#"
        strGraphs = "/" .join([strBaseDir,strPref,strFac,strGrDir])
        strFitsWDS = "/" .join([strGraphs,"tmp_FT.mdb"])

        numGraphs = ms_fac_grs[strFac][0]
        strFacTabs = ";" .join([strFac] * numGraphs)
        strFacList = ms_fac_grs[strFac][1]
        strStdTabs = ";" .join([strStd] * numGraphs)
        strTitle = ms_fac_grs[strFac][2]

        strNPopLats = ";" .join([strPopLats] * numGraphs)
        strNEpochWDS = ";" .join([strEpochWDS] * numGraphs)
        strNNormWDS = ";" .join([strNormWDS] * numGraphs)

        for (rper, lper, yr) in pers_yrs:
            strRPers = ";" .join([rper] * numGraphs)
            strLPers = ";" .join([lper] * numGraphs)
            strYrs = ";" .join([yr] * numGraphs)

# select epoch data
    if blnSelect:
        gp.AddMessage("Select: " + ";" .join([strFac, rper, lper]))
        SelectEpochData.main( strResDB, strPrefWDS, rper, strGridWDS, strUrbWDS,
                               yr, lper, strFacsWDS, strEpochWDS, strSelScratchWDS, "f" )
        for strLBin in lat_bins:
            strBinsList = ";" .join([strLBin] * numGraphs)

# process epoch data
        if blnProcess:
            gp.AddMessage(
                "Process: " + ";" .join([strPref,strFac,rper,lper,strLBin]))
            ProcessEpochLats.main(strEpochWDS,strFacsWDS,rper,lper,strPopLats,
                                   yr, strStdPopWDS, strLBin, strNormWDS, strScratchWDS, "f" )

# plot graphs
        for gu in gu_facs:
            strGU = ";" .join([gu] * numGraphs)
            if blnGraphs:
                gp.AddMessage(
                    "Plot: " + ";" .join([strFac,rper,lper,strStd,strLBin,gu]))
                PlotGraphsX4.main( strTitle, strGU, strFacTabs, strFacList,
                                   strNEpochWDS, strRPers, strLPers, strNPopLats, strYrs,
                                   strStdTabs, strBinsList, strNNormWDS, strFitsWDS, strGraphs,
                                   strScratchWDS, "f" )

### Done ###

```

PlotGraphsX4.py

```
# -----
# PlotGraphsX4.py
# -----
#
# NB Still uses central latitude - weighted centroid only in manual R code
#
# e.g.
# PlotGraphsX4 ....
#subtitle line
#aa;aa;aa;aa
#D:\GIS\EpochData.mdb
#prv;0a4;0a4;0a4
#prv;prv;0a4;prv
#D:\GIS\LatPops.mdb
#2006;2006;1956;2006
#euro;euro;euro;euro
#2.0;2.0;2.0;1.0
#D:\GIS\NormData.mdb
#D:\GIS\Fits.mdb
#D:\GIS\Png\Graphs
#D:\Workspace\Scratch.mdb f

# Import system modules
import sys, string, os, time, math, arcgisscripting

# move R to start of search path to fix lapack not loaded error
# from http://www.mail-archive.com/rpy-list@lists.sourceforge.net/msg01823.html

oldPath = os.environ['PATH'].split(os.pathsep)
import rpy2.robjects as robjects
newPath = os.environ['PATH'].split(os.pathsep)
newPath = os.pathsep.join(newPath[len(oldPath):] + newPath[:len(oldPath)])
# and the R objects
r = robjects.r
r('Sys.setenv(PATH="%s")' % newPath.replace('\\', '\\\\'))

from reportlab.pdfgen import canvas
from reportlab.lib.pagesizes import A4
from reportlab.lib.units import inch, cm
from reportlab.lib.utils import ImageReader
from reportlab.platypus import Table
from reportlab.lib import colors

import LatitudeFunctions as LF
import LatitudeDataNames as LDN
import PlotFunctions as PF

# Create the Geoprocessor object
gp = arcgisscripting.create(9.3)

def main(*argv):
    strCodeSec = "getting parameters"
    strExtraInfo = ""
    try:

        # Script arguments...
        strCmdArgs = ""
# nearly all the same as ProcessEpochLats since want to plot source
# data not just normalised, but most params changed to lists
        iParam=0
        in_SubTitle = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_SubTitle == '#': in_SubTitle = ""
        strCmdArgs = strCmdArgs + " " + in_SubTitle
        iParam=iParam+1

        in_GUFacs = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_GUFacs == '#': in_GUFacs = ""
        strCmdArgs = strCmdArgs + " " + in_GUFacs
        iParam=iParam+1
        in_FacTypes = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_FacTypes == '#': in_FacTypes = ""
        strCmdArgs = strCmdArgs + " " + in_FacTypes
        iParam=iParam+1
        in_Factors = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_Factors == '#': in_Factors = ""
        strCmdArgs = strCmdArgs + " " + in_Factors
        iParam=iParam+1

        in_EffWDS = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_EffWDS == '#': in_EffWDS = LDN.WDSMSLat()
        strCmdArgs = strCmdArgs + " " + in_EffWDS
        iParam=iParam+1

        in_ResPeriods = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_ResPeriods == '#': in_ResPeriods = ""
        strCmdArgs = strCmdArgs + " " + in_ResPeriods
        iParam=iParam+1
```

```

in_LifePeriods = argv[iParam] # gp.GetParameterAsText(iParam)
if in_LifePeriods == '#': in_LifePeriods = ""
strCmdArgs = strCmdArgs + " " + in_LifePeriods
iParam=iParam+1

in_LatPopWDS = argv[iParam] # gp.GetParameterAsText(iParam)
if in_LatPopWDS == '#': in_LatPopWDS = LDN.WDSLatPop()
strCmdArgs = strCmdArgs + " " + in_LatPopWDS
iParam=iParam+1

in_CYears = argv[iParam] # gp.GetParameterAsText(iParam)
if in_CYears == '#': in_CYears = ""
strCmdArgs = strCmdArgs + " " + in_CYears
iParam=iParam+1

in_StdPops = argv[iParam] # gp.GetParameterAsText(iParam)
if in_StdPops == '#': in_StdPops = ""
strCmdArgs = strCmdArgs + " " + in_StdPops
iParam=iParam+1

in_LatBins = argv[iParam] # gp.GetParameterAsText(iParam)
if in_LatBins == '#': in_LatBins = ""
strCmdArgs = strCmdArgs + " " + in_LatBins
iParam=iParam+1

in_MSNormWDS = argv[iParam] # gp.GetParameterAsText(iParam)
if in_MSNormWDS == '#': in_MSNormWDS = LDN.WDSMSNorm()
strCmdArgs = strCmdArgs + " " + in_MSNormWDS
iParam=iParam+1

in_StatsWDS = argv[iParam] # gp.GetParameterAsText(iParam)
if in_StatsWDS == '#': in_StatsWDS = LDN.WDSStats()
strCmdArgs = strCmdArgs + " " + in_StatsWDS
iParam=iParam+1

in_GraphWDS = argv[iParam] # gp.GetParameterAsText(iParam)
if in_GraphWDS == '#': in_GraphWDS = LDN.WDSGraph()
strCmdArgs = strCmdArgs + " " + in_GraphWDS
iParam=iParam+1

in_ScratchWDS = argv[iParam] # gp.GetParameterAsText(iParam)
if in_ScratchWDS == '#': in_ScratchWDS = LDN.WDSScratch()
strCmdArgs = strCmdArgs + " " + in_ScratchWDS
iParam=iParam+1

in_KTO = argv[iParam] # gp.GetParameterAsText(iParam)
if in_KTO == '#': in_KTO = ""
strCmdArgs = strCmdArgs + " " + in_KTO
iParam=iParam+1

blnKTO = (in_KTO.lower() == "true")

#gp.AddMessage("Called with effective args:\n%s" % strCmdArgs)

### Vars, Files and Fields ###

xName = ['MSnum', 'LatPop', 'MSpht', 'MSast']
yName = ["LatC_num", "LatC_Pop", "LatC_pht", "LatC_ast"]
yLUx = dict(zip(xName,yName))
# now have four pages output - almost the same so much code repeated
# - use loop but adds another list level onto output & data vars
NumPages = 4
pNums = range(NumPages)
pCode = ["std", "pct", "log", "lgt"]

# fit types for each page code
NumFits = 2
fNums = range(NumFits)
fCode = xName[2:] # ['MSpht', 'MSast']
strFTypeLab = { 'MSpht': ["Crude Incidence"], 'MSast': ["Age Standardised"] }
for ms in fCode:
    strFTypeLab[ms].append(strFTypeLab[ms][0] + " %")
    strFTypeLab[ms].append("Log " + strFTypeLab[ms][0])
    strFTypeLab[ms].append("Logit " + strFTypeLab[ms][0])

strFTypeFld = { 'MSpht': "crude", 'MSast': "agestd"}
strGXTabLab = ["AgeStd\nGraph", "AgeStd %\nGraph", "log(Std)\nGraph", "logit(Std)\nGraph"]

xDesc = [ ["MS Cases per Deg Lat", "Population (1000) per Deg Lat",
           "MS Cases per 100,000 Crude", "MS Cases per 100,000 Age Std."],
          ["MS Cases per Deg Lat", "Population (1000) per Deg Lat",
           "MS Cases per 100,000 % Crude", "MS Cases per 100,000 % Age Std."],
          ["MS Cases per Deg Lat", "Population (1000) per Deg Lat",
           "log(MS Cases Crude)", "log(MS Cases Age Std.)"],
          ["MS Cases per Deg Lat", "Population (1000) per Deg Lat",
           "logit(MS Cases Crude)", "logit(MS Cases Age Std.)"]]
fDesc = [ ["Fit to Crude Rate MS Cases", "Confidence Bands for Crude",
           "Fit to Age Std. MS Cases", "Confidence Bands for Age Std."],
          ["Fit to % Crude Rate MS Cases", "Confidence Bands for % Crude",
           "Fit to % Age Std. MS Cases", "Confidence Bands for % Age Std."],
          ["Log Fit to Crude Rate MS Cases", "Conf. Bands for Log Crude",
           "Log Fit to Age Std. MS Cases", "Conf. Bands for Log Age Std."],
          ["Logit Fit to Crude Rate MS Cases", "Conf. Bands for Logit Crude",
           "Logit Fit to Age Std. MS Cases", "Conf. Bands for Logit Age Std."]]

# need way of identifying page-wide files such as legend
# too many list params to use mangled params so use timestamp
# need seconds since script could be in batch

```

```

    strTimeStamp = time.strftime("%y%m%d%H%M%S")
# put files in sub-dir of in_GraphWDS, apart from main pfs
strGrDir = in_GraphWDS
strGrSubDir = in_GraphWDS + "/" + strTimeStamp
strCodeSec = "creating graph sub-dir"
strExtraInfo = "; %s" % strGrSubDir
os.mkdir(strGrSubDir)

# non-WDS params now lists
strCodeSec = "splitting param lists"
strExtraInfo = ""

strSubTitle = in_SubTitle
lstGUFac = in_GUFacs.split(";")
lstFacType = in_FacTypes.split(";")
lstFactor = in_Factors.split(";")
lstResPeriod = in_ResPeriods.split(";")
lstLifePeriod = in_LifePeriods.split(";")
lstCYear = in_CYears.split(";")
lstStdPop = in_StdPops.split(";")
lstLatBins = in_LatBins.split(";")
lstEffWDS = in_EffWDS.split(";")
lstLatPopWDS = in_LatPopWDS.split(";")
lstMSNormWDS = in_MSNormWDS.split(";")

AllLists = [lstFactor, lstFacType, lstEffWDS, lstLatPopWDS, lstMSNormWDS,
            lstGUFac, lstResPeriod, lstLifePeriod, lstCYear, lstStdPop, lstLatBins]
#if lstFactor and lstFacType:
blnNoFacs = ((lstFactor == ['']) and (lstFacType == ['']))
if blnNoFacs:
    lenLists = [len(x) for x in AllLists[2:]]
else:
    lenLists = [len(x) for x in AllLists]

nLists = set(lenLists)
if len(nLists) != 1:
    gp.AddMessage("Mismatch in param list lengths: %s" % repr(lenLists))
    gp.AddMessage("got blnNoFacs, lstFactor, lstFacType: %r\n%r\n%r" %
(bltnoFacs, lstFactor, lstFacType))
    strExtraInfo = "; keeping tool open"
    exit() # keep tool open
# nb: can't index set - need list here
numGraphs = lenLists[0]
maxGraphs = 4
if numGraphs > maxGraphs:
    gp.AddMessage("Max %d graphs at present" % maxGraphs)
    #return 0
    strExtraInfo = "; keeping tool open"
    exit() # keep tool open
GraphNums = range(numGraphs)
#if not lstFactor or not lstFacType:
if blnNoFacs:
    lstFacType = [""] * numGraphs
    lstFactor = [""] * numGraphs

strCodeSec = "parsing param lists"
strExtraInfo = ""

### Don't use "*" in place of outer list comprehension ###
### - get pointer to same list rather than multiple lists! ###
# (strRPer, strRPeriod, RPerYears, RPerMonths,
# strLPer, strLPeriod, LPerYears, LPerMonths,
# strPPer, strPPeriod, PPerYears, PPerMonths,
# strCYear, strGUFac, strFactor, strFacType
# ) = ([[None] * maxGraphs] for x in range(16))

# map - more condensed code; list comprehension - more efficient
# strFactor = map(LF.parseFactor, lstFactor)
# strCYear = map(LF.parseCYear, lstCYear)
# strStdPop = map(LF.parseStdPop, lstStdPop)
strGUFac = [LF.parseFactor(lstGUFac[n]) for n in GraphNums]
strFactor = [LF.parseFactor(lstFactor[n]) for n in GraphNums]
strFFacs = [strGUFac[n]+strFactor[n] for n in GraphNums]
strFacType = [LF.parseFacType(lstFacType[n]) for n in GraphNums]

strLabGUFac = LDN.AllGUFacLabels()
strLabFactor = [LDN.AllFactorLabels(strFacType[n]) for n in GraphNums]

strCYear = [LF.parseCYear(lstCYear[n]) for n in GraphNums]
strStdPop = [LF.parseStdPop(lstStdPop[n]) for n in GraphNums]

strCodeSec = "parsing periods"
in_PopPeriod = "all"
# for n in GraphNums:
#     strExtraInfo = "n = %d" % n
#     (strRPer[n], strRPeriod[n], RPerYears[n], RPerMonths[n])=LF.parsePeriod(lstResPeriod[n])
#     (strLPer[n], strLPeriod[n], LPerYears[n], LPerMonths[n])=LF.parsePeriod(lstLifePeriod[n])
#     (strPPer[n], strPPeriod[n], PPerYears[n], PPerMonths[n])=LF.parsePeriod(in_PopPeriod)
# above doesn't work, but illustrates aim of following!
rp = [LF.parsePeriod(lstResPeriod[n]) for n in GraphNums]
strRPer, strRPeriod, RPerYears, RPerMonths = [[rp[j][i] for j in GraphNums] for i in range(4)]

```



```

lp = [LF.parsePeriod(lstLifePeriod[n]) for n in GraphNums]
strLPer, strLPeriod, LPerYears, LPerMonths = [[lp[j][i] for j in GraphNums] for i in range(4)]
pp = [LF.parsePeriod(in_PopPeriod) for n in GraphNums]
strPPer, strPPeriod, PPerYears, PPerMonths = [[rp[j][i] for j in GraphNums] for i in range(4)]

strCodeSec = "setting up luts"
strExtraInfo = ""
dblLatBins = [float(lstLatBins[n]) for n in GraphNums]
strLatBins = [LF.parseLat(lstLatBins[n]) for n in GraphNums]
lutLB = dict([lstLatBins[i], i] for i in GraphNums)

strCodeSec = "setting tables"
strExtraInfo = "strBLatsTab"
# input tables of numbers of cases per latitude bin
strBLatsTabName = [LDN.KeepLatENums(strRPer[n], strLPer[n], strLatBins[n], strFactor[n])
                  for n in GraphNums]
strBLatsTab = [lstEffWDS[n] + "/" + strBLatsTabName[n] for n in GraphNums]

strExtraInfo = "strPLatsTab"
# input tables of census population per latitude bin
strPLatsTabName = [LDN.KeepLatPop(strCYear[n], strLatBins[n], strGUFac[n])
                  for n in GraphNums]
strPLatsTab = [lstLatPopWDS[n] + "/" + strPLatsTabName[n] for n in GraphNums]

strExtraInfo = "strNLatsTab"
# input tables of number of MS cases normalised by census population
strNLatsTabName = [LDN.KeepLatEPNorm(strRPer[n], strLPer[n], strCYear[n],
                                   strStdPop[n], strLatBins[n], strFactor[n]) for n in GraphNums]
strNLatsTab = [lstMSNormWDS[n] + "/" + strNLatsTabName[n] for n in GraphNums]

strExtraInfo = "strGraphLat"
# output graph of MS cases, census population and normalised
strGraphLatName = [LDN.KeepGraphNoExt(strGUFac[n], strFactor[n], strRPer[n], strLPer[n],
                                       strCYear[n], strStdPop[n], strLatBins[n]) for n in GraphNums]
# in_GraphWDS
strGraphPng = [{"%s/%s_%s.png" % (strGrSubDir, p, n) for n in strGraphLatName}
              for p in pCode ]
strGraphPdf = [{"%s/%s_%s.pdf" % (strGrSubDir, p, n) for n in strGraphLatName}
              for p in pCode ]

strExtraInfo = "strStatsTab"
strStatsTabN = LDN.KeepStatsX(strTimeStamp)
strStatsTabName = [strStatsTabN + "_" + p for p in pCode]
strStatsTab = [in_StatsWDS + "/" + n for n in strStatsTabName]
strStatsCSV = [strGrSubDir + "/" + n + ".csv" for n in strStatsTabName] # in_GraphWDS

strExtraInfo = "strXStatsTab"
strXStatsTabN = LDN.KeepCrossStatsX(strTimeStamp)
strXStatsTabName = [strXStatsTabN + "_" + p for p in pCode]
strXStatsTab = [in_StatsWDS + "/" + n for n in strXStatsTabName]
strXStatsCSV = [strGrSubDir + "/" + n + ".csv" for n in strXStatsTabName] # in_GraphWDS

strExtraInfo = "strTmpStatsTab"
strTmpStatsTabName = ["tmpFitStats_" + p for p in pCode]
strTmpStatsTab = [in_ScratchWDS + "/" + n for n in strTmpStatsTabName]

strExtraInfo = "strTmpXStatsTab"
strTmpXStatsTabName = ["tmpFitXStats_" + p for p in pCode]
strTmpXStatsTab = [in_ScratchWDS + "/" + n for n in strTmpXStatsTabName]

# and for all Graphs

strExtraInfo = "strGraphLeg"
strGraphLegName = LDN.KeepGraphLegX(strTimeStamp)
strGraphLeg = [{"%s/%s_%s" % (strGrSubDir, p, strGraphLegName) for p in pCode} # in_GraphWDS

strExtraInfo = "strPagePdf"
strPagePdfName = LDN.KeepGraphPdfX(strTimeStamp)
strPagePdf = [{"%s/%s_%s" % (strGrDir, p, strPagePdfName) for p in pCode} # in_GraphWDS

strCodeSec = "setting fields"
strExtraInfo = "FieldLats dicts"
# fields

strFld_lp = [LDN.AllFieldsLatPop(lstGUFac[n], lstCYear[n]) for n in GraphNums]
strFld_lno = [LDN.FieldLatNum(lstGUFac[n]) for n in GraphNums]
strFld_pht = [LDN.FieldLatNumPHT(lstGUFac[n]) for n in GraphNums]
strFld_ast = [LDN.FieldLatNumAST(lstGUFac[n]) for n in GraphNums]

strExtraInfo = "other fields"
strFld_lc = LDN.FieldLatC()
strFld_bin = "bin_size"
# factors
strFld_gen = LDN.FieldGender()
strFld_urb = LDN.FieldUrban()
strFld_fac = LDN.FieldFactor()
strFld_rpr = LDN.FieldResPer()
strFld_lpr = LDN.FieldLifePer()
strFld_cyr = LDN.FieldCYear()
strFld_std = LDN.FieldStdPop()
strFld_num = "sample_size"

```

```

strFld_grd, strFld_int, strFld_seg, strFld_sei, strFld_rsqr = [[] for i in range(5)]
for f in fNums:
    pref = strFTypeFld[fCode[f]]
    strFld_grd.append(pref + "_gradient")
    strFld_int.append(pref + "_intercept")
    strFld_seg.append(pref + "_se_grad")
    strFld_sei.append(pref + "_se_int")
    strFld_rsqr.append(pref + "_r_squared")

strFld_cmb1 = "combination_1"
strFld_cmb2 = "combination_2"
strFld_gno1 = "graph_num_1"
strFld_gno2 = "graph_num_2"
strFld_ftst = "f_test"
strFld_pval = "p_val"

# friendlier names for pdf tables
fieldlabs = {
    strFld_ftst: "F Test",
    strFld_pval: "P Val"
}
for i in fNums:
    fieldlabs[strFld_grd[i]] = "Gradient"
    fieldlabs[strFld_int[i]] = "Intercept"
    fieldlabs[strFld_seg[i]] = "StdErr Grad"
    fieldlabs[strFld_sei[i]] = "StdErr Int"
    fieldlabs[strFld_rsqr[i]] = "R Squared"

### Get Data ###
strCodeSec = "slurping data"
strExtraInfo = ""
# data files aren't huge so let's just get the lot now then deal with the plotting...

# dataFiles gives a list of the files we need to read data from
# The lat pop data is now split by gu into different data files
# Each item in the dataFiles list has a corresponding dict of
# required Y fields in the dataFldsY list, keys are the gu's
# If the dict value is a list of more than one Y field then the
# vals need to be summed, else just take the single val

dataFiles = [strBLatsTab, strPLatsTab, strNLatsTab, strNLatsTab]
dataFldsY = [strFld_lno, strFld_lp, strFld_pht, strFld_ast]

dataFldX = strFld_lc
numData = len(dataFiles)
nEno = 0
nPht = numData - 2
nAst = numData - 1
nDat = [nPht, nAst]
nPLo = 1
nPop = range(nPLo, nPht)
# do we want to scale by LatBin size
binFac = [1, 1, 0, 0]

# do we want to scale the data
dataFac = [0, 1000, 0, 0]

yType = [0, 1, 2, 3]
numYTypes = 4

gp.AddMessage("Reading in data...")
strCodeSec = "reading in data"
strExtraInfo = ""
Y = [[] for n in GraphNums] for i in range(numYTypes)]
Ys = [[] for n in GraphNums] for i in range(numYTypes)]
for d in range(0, numData):
    dFile = dataFiles[d]
    dataFldY = dataFldsY[d]

    for n in range(0, numGraphs):
        strExtraInfo = "%d %d" % (d, n)
        if binFac[d]:
            fac = 1.0 / float(lstLatBins[n])
        else:
            fac = 1.0
        if dataFac[d]:
            fac = fac / dataFac[d]

        strExtraInfo = "%d %d search cursor %s" % (d, n, dFile[n])
        get_rows = gp.SearchCursor(dFile[n], "", "", "", dataFldX + " A")
        get_row = get_rows.Next()

        if not get_row:
            gp.AddMessage("Cannot get data from %s" % dFile[n])
            strExtraInfo = "keeping tool open"
            exit() # keep tool open

# NA handling broken in 2.0 versions of RPy2
# replaced by NA_Real etc in 2.1 versions
# so instead of setting to NA equiv just don't add to lists

```

```

# however, must use matching LatC list for data
# whether different files or Y fields in same file
# note: with lat pops split across files, not all files have all gu's
# also: skipping NA/0 also good for log/logit!
while get_row:
    fldY = dataFldY[n]
    if isinstance(fldY,list):
        raw_val = 0
        for fY in fldY:
            strExtraInfo = "%d %d file %s; fY (%s) %s"%(d,n,dFile[n],type(fY),fY)
            raw_val += get_row.GetValue(fY)
    else:
        strExtraInfo = "%d %d file %s; fldY (%s) %s"%(d,n,dFile[n],type(fldY),fldY)
        raw_val = get_row.GetValue(fldY)
    if raw_val != 0:
        strExtraInfo = "%d %d file %s; raw_val %s * %s" % (d, n,
            dFile[n],repr(raw_val),repr(fac))

        val = raw_val * fac
        strExtraInfo = "%d %d file %s; fieldX %s" % (d, n, dFile[n], dataFldX)
        Xlat = get_row.GetValue(dataFldX)
        strExtraInfo = "%d %d file %s; Y %s"%(d,n,dFile[n],repr(Y[yType[d]][n]))
        if not Y[yType[d]][n].has_key(Xlat):
            Y[yType[d]][n][Xlat] = 0.0
            Ys[yType[d]][n][Xlat] = 0.0
            Y[yType[d]][n][Xlat] += val
            Ys[yType[d]][n][Xlat] += raw_val

        strExtraInfo = "%d %d file %s next" % (d, n, dFile[n])
        get_row = get_rows.Next()

# done with the cursor
del get_row
del get_rows

strExtraInfo = "calculating Ynum"
Ynum = [int(sum(Ys[0][n].values())) for n in GraphNums]
strYnum = [str(Ynum[n]) for n in GraphNums]

### Prep Data for R ###
# extent of pop data used for binning and normalising MS data
# so X's should all agree for given lat bin
# want graph for each lat bin showing counts, pop and norm
# want main graph showing all norms for different lat bins

strCodeSec = "setting up data"
strExtraInfo = ""
dataL = []

# save generated fits and stats as we go
lmFit = [{ fCode[0]: [{ } for n in GraphNums],
           fCode[1]: [{ } for n in GraphNums] } for p in pNums]

strExtraInfo = "; stats fields"
statFlds = [strFld_fac, strFld_rpr, strFld_lpr, strFld_cyr, strFld_std, strFld_bin]
for i in [0,1]:
    statFlds.extend([strFld_grd[i],strFld_int[i],strFld_seg[i],strFld_sei[i],strFld_rsq[i]])
otherFlds = statFlds[6:]
otherLabs = [fieldlabs[f] for f in otherFlds]

strExtraInfo = "; field formats"
# for table of stats use std 'g' format for all, but in-house format for rsq
fmtFlds = dict( [f, "g"] for f in statFlds )
for f in fNums: fmtFlds[strFld_rsq[f]] = "#"

strExtraInfo = "; initial stats data"
dblStatsData = [[dict([fld, None] for fld in otherFlds) for i in GraphNums] for p in pNums]
dblXStatsData = [None for p in pNums]

# so can refer to eg dblStatsData[p][n]['gradient']

HunThou = 100000.0
for n in GraphNums:
    strExtraInfo = "; dataL n:%d" % n
# - need these sorted into order - do keys and vals separately
dic = {
    'LatC_num': sorted(Y[yType[nEno]][n].keys()),
    'LatC_pop': sorted(Y[yType[nPLo]][n].keys()),
    'LatC_pht': sorted(Y[yType[nPht]][n].keys()),
    'LatC_ast': sorted(Y[yType[nAst]][n].keys())
}

dic['MSnum'] = [[Y[yType[nEno]][n][k] for k in dic[yLUx['MSnum']]]] #, [], []
dic['LatPop'] = [[Y[yType[nPLo]][n][k] for k in dic[yLUx['LatPop']]]] #, [], []
for i in pNums[1:]:
    dic['MSnum'].append(dic['MSnum'][0])
    dic['LatPop'].append(dic['LatPop'][0])
for i in fNums:
    dic[fCode[i]] = [[Y[yType[nDat[i]]][n][k] for k in dic[yLUx[fCode[i]]]]] #, [], []
    for j in pNums[1:]:
        dic[fCode[i]].append([])
# pct ([1]) moved to after first fit - now scaled by area under fit

```

```

# log & logit - shouldn't have any 0's - skipped on reading in data
dic[fCode[i]][2] = [math.log(y / HunThou) for y in dic[fCode[i]][0]]
dic[fCode[i]][3] = [math.log(y / (HunThou - y)) for y in dic[fCode[i]][0]]

idx = { 'LatC_num': dict(zip(dic['LatC_num'],range(len(dic['LatC_num']))))),
        'LatC_Pop': dict(zip(dic['LatC_Pop'],range(len(dic['LatC_Pop']))))})

strExtraInfo = "%d; setting dataL[n]\n%s" % (n, repr(dic))
dataL.append(dic)
strExtraInfo = "%d; dicMain dic" % n
### end first n in GraphNums loop
rName = { 'LatC_pht': "LatCN", 'LatC_ast': "LatCA",
          'MSpht': ["Npht%d" % p for p in pNums],
          'MSast': ["Nast%d" % p for p in pNums] }
# normalisation by area needs lat range of interest
LatRange = 14.0
LatMid = -41.0

gp.AddMessage("Doing Fits...")
# bit like graph labels below, but not quite
strXStatsLabs = map(";",".join,
                    zip(strFFacs,lstLatBins,strRPer,strLPer,strCYear,strStdPop,strYnum))

for p in pNums:
    gp.AddMessage(["...standard...","...percent...","...log...","...logit..."][p])
    strCodeSec = "doing stats"
    strExtraInfo = "; setting up stats table"
    strTab = strTmpStatsTab[p]
    try:
        if gp.Exists(strTab):
            gp.Delete(strTab)
    except:
        pass
    #gp.AddMessage("Error trying to check for and delete %s" % strTab)
    gp.CreateTable(in_ScratchWDS, strTmpStatsTabName[p])

#####
# this short bit is slowest (more than half) of whole script... why?
# - slow if use mdb for scratch - fast if use in_memory!
gp.AddField(strTab, strFld_fac, "text", "", "", 5) # note len
gp.AddField(strTab, strFld_rpr, "text", "", "", 5) # note len
gp.AddField(strTab, strFld_lpr, "text", "", "", 5) # note len
gp.AddField(strTab, strFld_cyr, "short")
gp.AddField(strTab, strFld_std, "text", "", "", 4) # note len
gp.AddField(strTab, strFld_num, "short")
for fld in statFlds[5:]:
    gp.AddField(strTab, fld, "double")

#####

# create insert cursor for new table
ins_rows = gp.InsertCursor(strTab)

for n in GraphNums:
    strExtraInfo = "; starting graph = %d" % n
    ### Fit Data ###
    # need to check number of data points with num > 0 - if < 2 then can't fit!
    # if assume have either 0 or unique values then can use set...
    # (NB: will have same number for pht and ast)
    dL = dataL[n] # dic
    fit = [None for f in fNums] # must have 'None' here in case no data
    nVals = len(set(dL[fCode[0]][p]) - set({}))
    if nVals > 1:
        # fitting rotated plot so swap x & y!
        for i in fNums:
            strExtraInfo = "; fitting %s n = %d" % (fCode[i],n)
            fit[i] = PF.RStatsFit(robjects, dL[yLUx[fCode[i]]], dL[fCode[i]][p],
                                rName[yLUx[fCode[i]]], rName[fCode[i]][p])
            strExtraInfo = "; saving fit %s n = %d" % (fCode[i],n)
        # save fits for later
        lmFit[p][fCode[i]][n] = fit[i]
        # normalise data by area under fit for pct data
        # if no fit then set data accordingly
        if p == 0:
            f_i = fit[i]['intc']
            f_g = fit[i]['grad']
            if f_i and f_g:
                A = LatRange * (f_i + LatMid * f_g)
                dL[fCode[i]][1] = [100.0 * y / A for y in dL[fCode[i]][0]]
            else:
                dL[fCode[i]][1] = [100.0/LatRange for y in dL[fCode[i]][0]]
        else:
            # not enough points to fit - fill data structure accordingly
            strExtraInfo = "; adding dummy data for n = %d" % n
            for i in fNums:
                strExtraInfo = "; adding dummy data for %s n = %d" % (fCode[i],n)
                lmFit[p][fCode[i]][n] = PF.DummyStatsFit(robjects,
                                                         dL[yLUx[fCode[i]]], dL[fCode[i]][p])
            # if no fit then set pct data accordingly
            if p == 0:
                dL[fCode[i]][1] = [100.0/LatRange for y in dL[fCode[i]][0]]

```

```

### Save Stats ###
# save stats - allowing for no fits
# NB - must use fit_ vars here for test, not lmFit unless use lmFit[fit]
strExtraInfo = "; saving stats for n = %d; filling dblStatsData" % n
dS = dblStatsData[p][n]
for i in fNums:
    f = fit[i]
    dS[strFld_grd[i]] = f and f['grad']
    dS[strFld_int[i]] = f and f['intc']
    dS[strFld_seg[i]] = f and f['se_grad']
    dS[strFld_sei[i]] = f and f['se_intc']
    dS[strFld_rsqr[i]] = f and f['rsqr']

strExtraInfo = "; saving stats for n = %d; creating row" % n
ins_row = ins_rows.NewRow()

strExtraInfo = "; saving stats for n = %d; setting fields" % n
ins_row.SetValue(strFld_fac, strFFacs[n])
ins_row.SetValue(strFld_rpr, lstResPeriod[n])
ins_row.SetValue(strFld_lpr, lstLifePeriod[n])
ins_row.SetValue(strFld_cyr, int(lstCYear[n]))
ins_row.SetValue(strFld_std, lstStdPop[n])
ins_row.SetValue(strFld_num, Ynum[n])
ins_row.SetValue(strFld_bin, dblLatBins[n])
strExtraInfo = "; saving stats for n = %d; setting other fields" % n
for fld in otherFlds:
    ins_row.SetValue(fld, dS[fld])

strExtraInfo = "; saving stats for n = %d; inserting row" % n
ins_rows.InsertRow(ins_row)
### end n in GraphNums loop

strExtraInfo = "; copying tmp stats table to %s" % strStatsTab[p]
try:
    if gp.Exists(strStatsTab[p]):
        gp.Delete(strStatsTab[p])
except:
    gp.AddMessage("Error trying to check for and delete %s" % strStatsTab[p])
gp.CopyRows_management(strTab, strStatsTab[p])
# csv version
strExtraInfo = "; copying tmp stats table to %s" % strStatsCSV[p]
f=open(strStatsCSV[p], 'w')
f.write(','.join([strFld_fac, strFld_rpr, strFld_lpr,
                 strFld_cyr, strFld_std, strFld_num, strFld_bin ]))
for fld in otherFlds:
    f.write( "," + fld )
f.write("\n")
for n in GraphNums:
    f.write(','.join([strFFacs[n], lstResPeriod[n], lstLifePeriod[n],
                    lstCYear[n], lstStdPop[n], str(Ynum[n]), lstLatBins[n] ]))
    for fld in otherFlds:
        f.write( ',' + str(dblStatsData[p][n][fld]) )
    f.write("\n")

f.close()

### Cross Stats ###

strExtraInfo = "; doing extra stats"
# var tests for age standardised fits for different param sets
dblXStatsData[p] = PF.RStatsCross(gp, robjects, lmFit[p][fCode[1]], lmFit[p][fCode[1]],
                                  {'test': 'p', 'ftst': strFld_ftst, 'pval': strFld_pval})

strExtraInfo = "; setting up xstats table"
strXTab = strTmpXStatsTab[p]
try:
    if gp.Exists(strXTab):
        gp.Delete(strXTab)
except:
    gp.AddMessage("Error trying to check for and delete %s" % strXTab)
gp.CreateTable(in_ScratchWDS, strTmpXStatsTabName[p])

gp.AddField(strXTab, strFld_gnol, "short")
gp.AddField(strXTab, strFld_gno2, "short")
gp.AddField(strXTab, strFld_cmb1, "text", "", "", 40) # note len
gp.AddField(strXTab, strFld_cmb2, "text", "", "", 40) # note len
# more efficient if use [1,2] instead of range but use range later for same loop
for i in range(2):
    strI = "_%d" % (i+1)
    gp.AddField(strXTab, strFld_fac + strI, "text", "", "", 5) # note len
    gp.AddField(strXTab, strFld_rpr + strI, "text", "", "", 5) # note len
    gp.AddField(strXTab, strFld_lpr + strI, "text", "", "", 5) # note len
    gp.AddField(strXTab, strFld_cyr + strI, "short")
    gp.AddField(strXTab, strFld_std + strI, "text", "", "", 4) # note len
    gp.AddField(strXTab, strFld_num + strI, "short")

gp.AddField(strXTab, strFld_ftst, "double")
gp.AddField(strXTab, strFld_pval, "double")

# create insert cursor for new table
ins_rows = gp.InsertCursor(strXTab)

```

```

for n1 in GraphNums:
    for n2 in GraphNums:
        if n1 != n2:
            strExtraInfo = "; saving xstats n1,n2 = %d,%d; creating row" % (n1,n2)
            ins_row = ins_rows.NewRow()

            strExtraInfo = "; saving stats n1,n2 = %d,%d; setting fields" % (n1,n2)
            ins_row.SetValue(strFld_gno1, n1+1)
            ins_row.SetValue(strFld_gno2, n2+1)
            ins_row.SetValue(strFld_cmb1, strXStatsLabs[n1])
            ins_row.SetValue(strFld_cmb2, strXStatsLabs[n2])
            ng = [n1, n2]
            for i in range(2):
                strI = "_%d" % (i+1)
                n = ng[i]
                ins_row.SetValue(strFld_fac + strI, strFFacs[n])
                ins_row.SetValue(strFld_rpr + strI, lstResPeriod[n])
                ins_row.SetValue(strFld_lpr + strI, lstLifePeriod[n])
                ins_row.SetValue(strFld_cyr + strI, int(lstCYear[n]))
                ins_row.SetValue(strFld_std + strI, lstStdPop[n])
                ins_row.SetValue(strFld_num + strI, Ynum[n])

            strExtraInfo = "; saving xstats n1,n2 = %d,%d; setting stat fields"%(n1,n2)
            ins_row.SetValue(strFld_ftst, dblXStatsData[p][n1][n2][strFld_ftst])
            ins_row.SetValue(strFld_pval, dblXStatsData[p][n1][n2][strFld_pval])

            strExtraInfo = "; saving xstats n1,n2 = %d,%d; inserting row" % (n1,n2)
            ins_rows.InsertRow(ins_row)

strExtraInfo = "; copying tmp xstats table to %s" % strXStatsTab[p]
try:
    if gp.Exists(strXStatsTab[p]):
        gp.Delete(strXStatsTab[p])
except:
    #gp.AddMessage("Error trying to check for and delete %s" % strXStatsTab[p])
    gp.CopyRows_management(strXTab, strXStatsTab[p])
# csv version
strExtraInfo = "; copying tmp xstats table to %s" % strXStatsCSV[p]
f=open(strXStatsCSV[p], 'w')
f.write(','.join([strFld_gno1,strFld_gno2,strFld_cmb1,strFld_cmb2]))
for i in range(2):
    strI = "_%d" % (i+1)
    strIc = strI + ","
    f.write(" " + strIc.join([strFld_fac, strFld_rpr, strFld_lpr,
                            strFld_cyr, strFld_std, strFld_num] ) + strI)
f.write(','.join(["", strFld_ftst, strFld_pval]))
f.write("\n")
for n1 in GraphNums:
    for n2 in GraphNums:
        if n1 != n2:
            f.write(','.join([str(n1+1), str(n2+1),
                              strXStatsLabs[n1], strXStatsLabs[n2] ] ))
            for i in range(2):
                n = ng[i]
                f.write( ','.join(["", strFFacs[n], lstResPeriod[n],
                                    lstLifePeriod[n], lstCYear[n], lstStdPop[n], str(Ynum[n]) ] ))
            dn12 = dblXStatsData[p][n1][n2]
            f.write(','.join(["", str(dn12[strFld_ftst]), str(dn12[strFld_pval])]))
            f.write("\n")

f.close()

### end first p in pNums loop

gp.AddMessage("Plotting graphs...")
strCodeSec = "plotting data"
strExtraInfo = ""

# plotting code here
### Setup Graph Info ###
# do most here - some needed for embedded graphs, some for pdf file
# individual graph labels
strGraphLabs = map(" ".join,zip(strFFacs,lstLatBins,strRPer,strLPer,strCYear,strStdPop))
strGraphFacs = [strLabGUFac[strGUFac[n]] + "; " + strLabFactor[n][strFactor[n]]
                for n in GraphNums]

# combined graphs page label
strGraphCombLab = "; ".join(map(":".join,
                                zip([repr(n+1) for n in GraphNums], map(lambda x: "("+x+")", strGraphLabs))))
# combined graphs file subject
strGraphCombSub = "\n".join(map(lambda x: "("+x+")", strGraphLabs))

strPageHead = [ ["Fit Statistics for Crude Incidence and Age Standardised MS Case data",
                "(Factors, Latitude Bin, NZ Residence, Epochs, Pop Census Year, Std Pop):",
                strGraphCombLab, strSubTitle] for p in pNums]
strPageHead[0][0] = strPageHead[0][0] + " for"
strPageHead[1][0] = strPageHead[1][0] + " (% of Std Sample per Deg Lat) for"
strPageHead[2][0] = "Log " + strPageHead[0][0] # note use updated [0]
strPageHead[3][0] = "Logit " + strPageHead[0][0] # note use updated [0]

strPdfT1 = [ "Age Std graphs & fit stats", "Age Std graphs & fit stats as %",

```

```

                "Age Std graphs & log fit stats", "Age Std graphs & logit fit stats"]
strPdfTitle = [ strPdfT1[p] + ";" + strSubTitle for p in pNums ]
strPdfSubject = "Sets:\n" + strGraphCombSub
strPdfAuthor = "Diane Fraser"

strExtraInfo = "; setting offsets and sizes(1)"
x1 = 1.25 * cm
x2 = 11.25 * cm
x3 = x2
y1 = 0.5 * cm
y2 = 10.25 * cm
y3 = 20.0 * cm
imOffset = {'x': [x1, x2, x1, x2, x3],
            'y': [y2, y2, y1, y1, y3] }
imSize = {'x': 9 * cm, 'y': 9 * cm}

strLabMSDegLat = "MS Nums per Deg Lat"
strLabMSPHT = [{"MS Nums per 100,000 Pop"}, {"MS Nums % per Deg Lat"},
               ["log(MS Rate)", ["logit(MS Rate)"]]
strLabPopDegLat = "Pop (1000) per Deg Lat"
strLabLat = "Latitude"

strExtraInfo = "; setting offsets and sizes(2)"
# NB: rotated plots
pltMars = [(5,5,5,0), (5,0,5,5)]
pltVars = [['LatPop', 'MSnum'], fCode]
pltAxes = [[2,3], [1]]
pltLabs = [ {'ax': [1,2,3], 'ln': [3,3,3],
            'tx': [strLabPopDegLat, strLabLat, strLabMSDegLat]},
           {'ax': [1,3,3], 'ln': [3,1,3], 'tx': [""] * 3} ]
minX = dict([xName, [0]*NumPages ] for xName in yLUx.keys())
# log/it
maxX = {'MSnum': [800]*NumPages, 'LatPop': [2000]*NumPages,
        fCode[0]: [200,20,-4,-4], fCode[1]: [200,20,-4,-4]}
minX[fCode[0]][2] = -16
minX[fCode[1]][2] = -16
minX[fCode[0]][3] = -16
minX[fCode[1]][3] = -16
minY = -48.0
maxY = -34.0
pngX = 800
pngY = 800
legX = 500 # 600
legY = 400 # 800
xDiff = dict([ms, [maxX[ms][p] - minX[ms][p] for p in pNums]] for ms in fCode)
xStatP = dict([ms, [ [x * xDiff[ms][p] + minX[ms][p] for x in [0.60,0.625,0.625, 0.625] ]
                    for p in pNums]] for ms in fCode)
# if we add the sample size at the end we use the last fCode (MSast) xStat vals
xStatP['MSsmp'] = [xStatP[fCode[-1]][p][1] for p in pNums]
yD = 0.4
yTop = -34.0
yStatP = {}
#yStatP[fCode[0]] = [yTop - n * yD for n in range(4)]
#yTop = yStatP[fCode[0]][-1] - yD
#yStatP[fCode[1]] = [yTop - n * yD for n in range(4)]
#yTop = yStatP[fCode[1]][-1] - yD
NStats = 4
for f in fCode:
    yStatP[f] = [yTop - n * yD for n in range(NStats)]
    yTop = yStatP[f][-1] - yD
yStatP['MSsmp'] = [yTop]

strExtraInfo = "; setting offsets and sizes(3)"
data_cols = {'MSnum': "darkolivegreen", 'LatPop': "orange",
             fCode[0]: "Blue", fCode[1]: "darkgreen"}
fit_cols = {fCode[0]: {'fit': "purple", 'band': "magenta"},
            fCode[1]: {'fit': "cyan3", 'band': "aquamarine3"} }

f_pts_head = 10 # for header
f_pts_num = 7 # for graph num
f_cex_mtext = {'png': 1.6, 'pdf': 1} # for axes' labels
f_cex_axis = {'png': 2, 'pdf': 1} # for axes' scale
f_cex_stats = {'png': 1.6, 'pdf': 1} # for stats summary

leglabs = [[] for p in pNums]
for p in pNums:
    leglabs[p] = xDesc[p][:]
    leglabs[p].extend(fDesc[p][:])
legcols = [data_cols[x] for x in xName]
leglws = [3 for x in xName]
legltys = [1 for x in xName]
legpchs = [1 for x in xName]
for f in fCode:
    legcols.extend([fit_cols[f]['fit'], fit_cols[f]['band']])
    leglws.extend([3, 3])
    legltys.extend([1, 2])
    legpchs.extend([-1, -1])

leginst = [0.01, 0.005]

strExtraInfo = "; about to loop over pages and graphs"

```

```

### Plot Data ###
try:
    r['dev.off']() # in case anything left open...
except:
    pass

#gp.AddMessage("Creating pages...")
for p in pNums:
    # one file per param set with raw data and normalised

    for n in GraphNums:
        strExtraInfo = "; starting graph for p,n = %d, %d" % (p,n)
        dL = dataL[n]
        #fit_pht = lmFit[p]['Crude'][n]
        #fit_ast = lmFit[p]['AgeStd'][n]
        fit = dict((ms,lmFit[p][ms][n]) for ms in fCode)
        strGLab1 = strGraphLabs[n]
        strGLab2 = strGraphFacs[n]
    # want a png version for embedding in the main pdf page
    # plus a pdf version suitable for use in Illustrator
    for gdev in ['png', 'pdf']:
        if gdev == 'png':
            strExtraInfo = "; opening device n = %d, file = %s"%(n,strGraphPng[p][n])
            r.png(filename=strGraphPng[p][n], width=pngX, height=pngY)
        elif gdev == 'pdf':
            strExtraInfo = "; opening device n = %d, file = %s"%(n,strGraphPdf[p][n])
            r.pdf(file=strGraphPdf[p][n], paper="a4",
                title=strPdfTl[p]+"; " +strGraphLabs[n])
        else:
            pass
        strExtraInfo = "; setting layout p,n = %d %d %s" % (p, n, gdev)
        kwargs = {'byrow': True}
        r.layout(r.matrix(robjcts.IntVector([1,2,2]), 1, 3, **kwargs))

    # two graphs back to back
    for plt in [0,1]:
        pltLabs[plt]['sz'] = [f_cex_mtext[gdev]]*3
        pltLabs[1]['tx'] = [strLabMSPHT[p], strGLab1, strGLab2]

    for plt in [0, 1]:
        strExtraInfo = "; constructing plot %d p,n = %d %d %s"%(plt, p, n, gdev)
        x0 = pltVars[plt][0]
        x1 = pltVars[plt][1]
        y0 = yLUx[x0]
        y1 = yLUx[x1]
        strExtraInfo += " with x0,x1,y0,y1 = %r %r %r %r" % (x0, x1, y0, y1)
        PF.AddPlot2(gp, robjcts, pltMars[plt], f_cex_axis[gdev],
            [dL[x0][p],dL[x1][p]], [dL[y0],dL[y1]],
            [[minX[x0][p], maxX[x0][p]], [minX[x1][p], maxX[x1][p]]],
            [[minY, maxY],[minY, maxY]], [1,1], [3,3],
            [data_cols[x0],data_cols[x1]])
        strExtraInfo = "; adding axes to plot %d p,n = %d %d %s"%(plt, p, n, gdev)
        for a in pltAxes[plt]: r.axis(a)
        strExtraInfo = "; adding labels to plot %d p,n = %d %d %s"%(plt, p, n, gdev)
        pL = pltLabs[plt]
        PF.AddMarLabels(robjcts, pL['ax'], pL['ln'], pL['sz'], pL['tx'])

### Plot Fits & Stats ###
strExtraInfo = "; adding fits to plot for n = %d %s" % (n, gdev)
for ms in fCode:
    PF.AddRFitRot(robjcts, fit[ms],
        [3,3], [1,2], [fit_cols[ms]['fit'],fit_cols[ms]['band']],
        xStatP[ms][p], yStatP[ms], f_cex_stats[gdev],
        [strFTypeLab[ms][p] + ":",
        "Gradient = %s (se:%s)" %
        (PF.Fmt(fit[ms]['grad'], "g",8,5), PF.Fmt(fit[ms]['se_grad'], "g",6,3)),
        "Intercept = %s (se:%s)" %
        (PF.Fmt(fit[ms]['intc'], "g",8,5), PF.Fmt(fit[ms]['se_intc'], "g",6,3)),
        "Corr.Coeff. = %s" % PF.Fmt(fit[ms]['rsq'], "#",5,4)])
    PF.AddTextRot(robjcts, xStatP['MSsmp'][p], yStatP['MSsmp'], f_cex_stats[gdev],
        ["Sample Size: %r" % Ynum[n]])

r['dev.off']()

### Legend ###
# one file for all graphs
PF.PlotLegendPng(robjcts, strGraphLeg[p], legX, legY, "Key for all Graphs",
    leglabs[p], legcols, leglws, legltys, legpchs, leginst)

### Setup page Pdf ###
strExtraInfo = "; putting graphs together"

strFile = strPagePdf[p] # .replace('\', '/')
strExtraInfo = "; getting canvas for file = %s" % strFile
canv = canvas.Canvas(strFile) # , pagesize=A4
strExtraInfo = "; setting page title"
canv.setTitle(strPdfTitle[p])
canv.setSubject(strPdfSubject)
canv.setAuthor(strPdfAuthor)
strExtraInfo = "; setting page compression"

```



```

    canv.setPageCompression(0)
### Load Graphs ###
for n in GraphNums:
    strExtraInfo = "; reading image n = %d, file = %s" % (n, strGraphPng[p][n])
    PF.LoadGraphFile(gp, canv, strGraphPng[p][n], imOffset['x'][n], imOffset['y'][n],
                    imSize['x'], imSize['y'])
    PF.GraphNum(canv, imOffset['x'][n], imOffset['y'][n]+imSize['y'], n+1, f_pts_num)

    strExtraInfo = "; loaded graphs"
### Load Legend ###
strExtraInfo = "; reading legend file = %s" % strGraphLeg[p]
imSizeX = imSize['x'] * legX / pngX
imSizeY = imSize['y'] * legY / pngY
imOffsetX = imOffset['x'][4] + imSize['x'] - imSizeX
PF.LoadGraphFile(gp, canv, strGraphLeg[p], imOffsetX, imOffset['y'][4],
                 imSizeX, imSizeY)
### Header ###
strExtraInfo = "; adding header for file = %s" % strFile
# add a header
yD = 0.5 * cm
yPosn = PF.PageHeader(canv, (1.0 * cm), (28.5 * cm), yD, strPageHead[p], f_pts_head)

### Stats Table ###
# and tables of stats
# first the fits
graph_tags = {'graph': "Graph", 'gname': [repr(n+1) for n in GraphNums],
              'fit1': strFTypeLab[fCode[0]][p], 'fit2': strFTypeLab[fCode[1]][p],
              'fits': otherLabs}
strExtraInfo = "; adding stats to graph for p, n = %d, %d" % (p,n)
yTab = PF.StdStatsTabFit2(canv, numGraphs, (1.0 * cm), yPosn,
                          graph_tags, dblStatsData[p], otherFlds, fmtFlds)

### Cross Stats ###
# now the f-tests comparing fits for different graphs
strExtraInfo = "; adding cross-stats for file = %s" % strFile
PF.CrossStatsTab(canv, numGraphs, (1.0 * cm), yTab - yD,
                 {'graph': strGXTabLab[p], 'gname': [repr(n+1) for n in GraphNums],
                  'test': fieldlabs[strFld_ftst], 'val': fieldlabs[strFld_pval]},
                 dblXStatsData[p], {'test': strFld_ftst, 'val': strFld_pval})
    canv.showPage()
    canv.save()
### end p in pNums loop

### Done ###
gp.AddMessage("Done.")
strCodeSec = "End!"
strExtraInfo = ""

if blnKTO:
    strCodeSec = "exception to keep tool open!"
    exit() # keep tool open in case works and gets to here!

except:
    gp.AddMessage("Got exception in PlotGraphsX4, section = %s %s" % (strCodeSec, strExtraInfo))
    raise
else:
    return 0 # exit errorlessly

if __name__ == '__main__':
    main(*sys.argv[1:])

```

PlotFunctions.py

```
# -----
# PlotFunctions.py
# -----
# functions for repetative plot stuff

# defines:
#

# Import system modules
import sys, string, os, math

from reportlab.lib.units import cm
from reportlab.lib.utils import ImageReader
from reportlab.platypus import Table
from reportlab.lib import colors

##### new X4 stuff #####

def AddPlot2(gp, robj, mar, caxis, xdata, ydata, xlim, ylim, lty, lwd, col):
    r = robj.r
    rMar = robj.FloatVector(mar)
    kwargs = {'new': False, 'mar': rMar, 'cex.axis': caxis}
    r.par(**kwargs)

    for i in range(len(xdata)):
        if len(xdata[i]) != len(ydata[i]):
            gp.AddMessage("Data length mismatch!\n - got lengths %d and %d:\n%s\n%s"
                % (len(xdata[i]), len(ydata[i]), xdata[i], ydata[i]))

            rxlimi = robj.FloatVector(xlim[i])
            rylimi = robj.FloatVector(ylim[i])
            kwargs = {'type': "o", 'lty': lty[i], 'lwd': lwd[i], 'col': col[i],
                'x': xdata[i], 'y': ydata[i], 'xlim': rxlimi, 'ylim': rylimi,
                'xlab': "", 'ylab': ""}

            if i == 0:
                kwargs['yaxt'] = "n"
            else:
                kwargs['axes'] = False

            r.plot(**kwargs)

            kwargs = {'new': True}
            r.par(**kwargs) # next graph won't clean the prev

    kwargs = {'new': False}
    r.par(**kwargs) # reset to false

def AddMarLabels(robj, side, line, cex, text):
    r = robj.r
    for i in range(len(text)):
        if text[i]:
            pargs = [text[i]]
            kwargs = {'side': side[i], 'line': line[i], 'cex': cex[i]}
            r.mtext(*pargs, **kwargs)

def RStatsFit(robj, xData, yData, xName, yName):
    # not weighted
    r = robj.r
    rX = robj.FloatVector(xData)
    rY = robj.FloatVector(yData)
    robj.globalEnv[xName] = rX
    robj.globalEnv[yName] = rY
    rFit = r.lm(yName + " ~ " + xName)

    rSumm = r.summary(rFit)
    summ = dict(zip(rSumm.names, rSumm))
    rsq = summ['r.squared'][0]

    kwargs = {'level': 0.95}
    rConf = r.confint(rFit, **kwargs)
    kwargs = {'interval': "confidence"}
    rPred = r.predict(rFit, **kwargs)

    results = dict(zip(rFit.names, rFit))
    rCoeffs = results['coefficients']
    coeffs = dict(zip(rCoeffs.names, rCoeffs))
    rDiag = r['ls.diag'](rFit)
    diag = dict(zip(rDiag.names, rDiag))
    stderr = dict(zip(diag['std.err'].names[0], diag['std.err']))

    grad = coeffs[xName]
    intc = coeffs['(Intercept)']
    se_grad = stderr[xName]
    se_intc = stderr['(Intercept)']
```

```

numPts = len(rX)
fhat = [rPred[i] for i in range(0,numPts)]
fhLb = [rPred[i] for i in range(numPts,2*numPts)]
fhUb = [rPred[i] for i in range(2*numPts,3*numPts)]

dic = {'rx': rX, 'ry': rY, 'fit': rFit, 'conf': rConf, 'pred': rPred,
       'grad': grad, 'intc': intc, 'se_grad': se_grad, 'se_intc': se_intc,
       'rsq': rsq, 'ypred': fhat, 'yconfl': fhLb, 'yconfu': fhUb}
return dic

def DummyStatsFit(robject, xData, yData):
    r = robject.r
    rX = robject.FloatVector(xData)
    rY = robject.FloatVector(yData)
    dic = {'rx': rX, 'ry': rY, 'fit': None, 'conf': None, 'pred': None,
          'grad': None, 'intc': None, 'se_grad': None, 'se_intc': None,
          'rsq': None, 'ypred': None, 'yconfl': None, 'yconfu': None}
    return dic

def RStatsCross(gp, robject, rStatsFit1, rStatsFit2, fld_keys):
    r = robject.r
    dblXStatsData = []
    blnList1 = isinstance(rStatsFit1,list)
    if blnList1:
        num1 = len(rStatsFit1)
    else:
        num1 = 1
        fit1 = rStatsFit1['fit']
        rX1 = rStatsFit1['rx']
        rY1 = rStatsFit1['ry']
    blnList2 = isinstance(rStatsFit2,list)
    if blnList2:
        num2 = len(rStatsFit2)
    else:
        num2 = 1
        fit2 = rStatsFit2['fit']
        rX2 = rStatsFit2['rx']
        rY2 = rStatsFit2['ry']
    strFld_ftst = "ftst"
    strFld_pval = "pval"
    str_test = "c"
    if isinstance(fld_keys, dict):
        if fld_keys.has_key("ftst"):
            strFld_ftst = fld_keys["ftst"]
        if fld_keys.has_key("pval"):
            strFld_pval = fld_keys["pval"]
        if fld_keys.has_key("test"):
            str_test = fld_keys["test"][0].lower()
    if str_test == "c":
        blnCoin = True
    else:
        blnCoin = False

    for n1 in range(0, num1):
        dblXStatsData.append([])
        if blnList1:
            fit1 = rStatsFit1[n1]['fit']
            rX1 = rStatsFit1[n1]['rx']
            rY1 = rStatsFit1[n1]['ry']
        for n2 in range(0, num2):
            dic = {}
            if blnList2:
                fit2 = rStatsFit2[n2]['fit']
                rX2 = rStatsFit2[n2]['rx']
                rY2 = rStatsFit2[n2]['ry']
    # still test fit1 and fit2
    # - if didn't fit either before then joint fit here not reasonable either
    if fit1 and fit2:
        pX = [x for x in rX1]
        pX.extend([x for x in rX2])
        pY = [y for y in rY1]
        pY.extend([y for y in rY2])
        pF = [0] * len(rX1)
        pF.extend([1] * len(rX2))
        rX = robject.FloatVector(pX)
        rY = robject.FloatVector(pY)
        rF = robject.IntVector(pF)

        xName = 'xtest'
        yName = 'ytest'
        fName = 'fac'
        robject.globalEnv[xName] = rX
        robject.globalEnv[yName] = rY
        robject.globalEnv[fName] = rF

        try:
            strModel = "%s ~ %s * factor(%s)" % (yName,xName,fName)
            rFit = r.lm(strModel)

            rAn = r.anova(rFit)

```

```

        anova = dict(zip(rAn.names, rAn))
        sumsq = anova['Sum Sq']
        degfr = anova['Df']

        # degfr is RVector - [1:-2] doesn't work!
        if blnCoin:
            num_df = sum([degfr[n] for n in range(1, len(degfr)-1)])
            num = sum([sumsq[n] for n in range(1, len(sumsq)-1)]) / num_df
        else: # just test for parallel
            num_df = degfr[len(degfr)-2] # should be 1
            num = sumsq[len(sumsq)-2] / num_df
            den_df = degfr[-1]

        den = sumsq[-1]/den_df

        F_test = num/den

        # can use vectors here for one set:
        p_val = r.df(F_test, num_df, den_df)[0]

        dic[strFld_ftst] = F_test
        dic[strFld_pval] = p_val
    except:
        dic[strFld_ftst] = None
        dic[strFld_pval] = None
    else:
        dic[strFld_ftst] = None
        dic[strFld_pval] = None
        dblXStatsData[n1].append(dic)
if blnList1 or blnList2:
    return dblXStatsData
else:
    return dic

def RStatsCrossX(gp, robj, rStatsFit1, rStatsFit2, fld_keys):
    r = robj.r
    dblXStatsData = []
    blnList1 = isinstance(rStatsFit1, list)
    if blnList1:
        num1 = len(rStatsFit1)
    else:
        num1 = 1
        fit1 = rStatsFit1['fit']
    blnList2 = isinstance(rStatsFit2, list)
    if blnList2:
        num2 = len(rStatsFit2)
    else:
        num2 = 1
        fit2 = rStatsFit2['fit']
    strFld_ftst = "ftst"
    strFld_pval = "pval"
    if isinstance(fld_keys, dict):
        if fld_keys.has_key("ftst"):
            strFld_ftst = fld_keys['ftst']
        if fld_keys.has_key("pval"):
            strFld_pval = fld_keys['pval']
    for n1 in range(0, num1):
        dblXStatsData.append([])
        if blnList1:
            fit1 = rStatsFit1[n1]['fit']
        for n2 in range(0, num2):
            dic = {}
            if blnList2:
                fit2 = rStatsFit2[n2]['fit']
            if fit1 and fit2:
                rVt = r['var.test'](fit1, fit2)
                vt = dict(zip(rVt.names, rVt))
                # 'statistic' 'parameter' 'p.value' 'conf.int' 'estimate'
                # 'null.value' 'alternative' 'method' 'data.name'
                stats = dict(zip(vt['statistic'].names, vt['statistic']))
                test_name = stats.keys()[0]
                test_val = stats[test_name]
                # test_confint = vt['conf.int']
                # conflev = r.attr(vt['conf.int'], 'conf.level')[0]
                test_pval = vt['p.value'][0]
                dic[strFld_ftst] = test_val
                dic[strFld_pval] = test_pval
            else:
                dic[strFld_ftst] = None
                dic[strFld_pval] = None
                dblXStatsData[n1].append(dic)
    if blnList1 or blnList2:
        return dblXStatsData
    else:
        return dic

def AddrFitRot(robj, rStatsFit, lwd, lty, col, xTxt, yTxt, fcex, lTxt):
    # adds fits and conf bands to plot with optional stats text
    # note rotated - x & y swapped

```

```

r = robj.r
rX = rStatsFit['rx']
if rStatsFit['ypred']:
    kwargs = {'y': rX, 'x': rStatsFit['ypred'], 'lwd': lwd[0], 'lty': lty[0], 'col': col[0]}
    r.lines(**kwargs)
if rStatsFit['yconfl']:
    kwargs = {'y': rX, 'x': rStatsFit['yconfl'], 'lwd': lwd[1], 'lty': lty[1], 'col': col[1]}
    r.lines(**kwargs)
if rStatsFit['yconfu']:
    kwargs = {'y': rX, 'x': rStatsFit['yconfu'], 'lwd': lwd[1], 'lty': lty[1], 'col': col[1]}
    r.lines(**kwargs)
if lTtxt:
    xtext = robj.FloatVector(xTtxt)
    ytext = robj.FloatVector(yTtxt)
    atext = robj.IntVector((0,0))
    ltext = robj.StrVector(lTtxt)
    kwargs = {'x': xtext, 'y': ytext, 'labels': ltext, 'adj': atext, 'cex': fcex}
    r.text(**kwargs)

def AddTextRot(robj, xTtxt, yTtxt, fcex, lTtxt):
# adds stats text
# note rotated - x & y swapped
    r = robj.r
    if lTtxt:
        xtext = robj.FloatVector(xTtxt)
        ytext = robj.FloatVector(yTtxt)
        atext = robj.IntVector((0,0))
        ltext = robj.StrVector(lTtxt)
        kwargs = {'x': xtext, 'y': ytext, 'labels': ltext, 'adj': atext, 'cex': fcex}
        r.text(**kwargs)

def PlotLegendPng(robj, strFile, xW, yH, title, labs, cols, lwds, ltys, pchs, inst):
    r = robj.r
    r.png(strFile, width=xW, height=yH)
    kwargs = {'byrow': True}
    r.layout(r.matrix(robj.IntVector([1]), 1, 1, **kwargs))
    X = []
    Y = []
    xlim = robj.FloatVector((0, 75))
    ylim = robj.FloatVector((0, 100))
    rlabs = robj.StrVector(labs)
    rcols = robj.StrVector(cols)
    rlwds = robj.IntVector(lwds)
    rltys = robj.IntVector(ltys)
    rpchs = robj.IntVector(pchs)
    rinst = robj.FloatVector(inst)
    bNew = False
    bAxes = False
    mar = robj.FloatVector((0,5,3,4)) # (0,5,5,4)
    kwargs = {'new': bNew, 'mar': mar}
    r.par(**kwargs)

    kwargs = {'type': "p", 'col': 0, 'axes': bAxes, 'xlab': "", 'ylab': "",
              'x': X, 'y': Y, 'xlim': xlim, 'ylim': ylim}
    r.plot(**kwargs)

    kwargs = {'title': title, 'cex': 1.6, 'pt.cex': 1.0, 'y.intersp': 1.2,
              'x': "bottomright", 'inset': rinst, 'legend': rlabs,
              'col': rcols, 'lwd': rlwds, 'lty': rltys, 'pch': rpchs}
    r.legend(**kwargs)

    r['dev.off']()

def LoadGraphFile(gp, canv, strFile, xPosn, yPosn, xSize, ySize):
    irGr = ImageReader(strFile)
    canv.drawImage(irGr, xPosn, yPosn, xSize, ySize)

def PageHeader(canv, xHead, yHead, yD, strHead, pts):
    canv.setFontSize(pts)
    for i in range(len(strHead)):
        canv.drawString(xHead, yHead - i * yD, strHead[i])
    canv.setFontSize(12)
    return yHead - len(strHead) * yD

def GraphNum(canv, xPosn, yPosn, num, pts):
    canv.setFontSize(pts)
    canv.drawString(xPosn, yPosn, str(num))
    canv.setFontSize(12)

# Python string formatting - note difference between e & g formats:
# e/E: The precision determines
# the number of digits after the decimal point and defaults to 6.
# g/G: The precision determines
# the number of significant digits before and after the decimal point and defaults to 6.

```

```

def Fmt(fNum, fmtReq, maxStrLen, maxSigDig, noNum=""):
# bit of overhead having this in function but getting messy in situ
# fmtReq - tells us the type of formatting we want:
# g: mimic g but be more efficient with the space available
# #: mixed - don't want exp for small numbers but don't lose info if large
# width is just minimum width and pad with space if necessary - don't need here
# if num == 0 then log10 errors - if very small but != 0 then okay
# so can just check == 0 rather than size of exponent
# first check for 'None' and return noNum as string
    if fNum is None:
        return str(noNum)

# places used by exp, if required
# assume 2 digit exp - adjust later if 3!
    lenExp = 4
    df = fNum
    adf = abs(df)
    sgn = abs(df < 0)
    maxDigLen = maxStrLen - 1 - sgn
    if maxDigLen < 1: maxDigLen = 1
    if maxSigDig < 1: maxSigDig = 1
    maxStdSigDig = min(maxSigDig, maxDigLen)
    maxExpSigDig = min(maxSigDig, maxDigLen - lenExp)
    if maxExpSigDig < 1: maxExpSigDig = 1
    precStd = maxStdSigDig - 1
    precExp = maxExpSigDig - 1

    if adf > 0:
        try:
            exp = int(math.floor(math.log10(adf)))
        except:
            exp = None
    else:
        exp = -1
    if exp is not None:
        if abs(exp) > 99 : precExp = precExp - 1
        if exp > precStd or (exp < -4 and fmtReq == "g"):
            fmt = "%%.%de" % precExp
        else:
            if exp > -1: precStd = precStd - exp - 1
            if precStd < 0: precStd = 0
            fmt = "%%.%df" % precStd
        try:
            fNum = (fmt % fNum).replace("e+0", "e+").replace("e-0", "e-")
        except:
            fNum = "Error:\nfmt = " + fmt + ("\nnum = %r" % fNum)
    else:
        # don't know what went wrong with log10 so just give us the number
        fNum = "%r" % fNum

    return fNum

def StdStatsTabFit2(canv, numGraphs, xLeft, yTop, graph_tags, graph_data, fit_keys, fmt_keys):
# the fits and f-tests for each graph

    stats_data = [[""] * (len(fit_keys) + 1)]
    stats_data[0][0] = graph_tags['graph']
    stats_data[0][1] = graph_tags['fit1']
    stats_data[0][1+len(fit_keys)/2] = graph_tags['fit2']

    stats_data.append([""])
    stats_data[1].extend(graph_tags['fits']) # fit_keys
    sigfig = 6
    strlen = 8
    for n in range(numGraphs):
        row = [graph_tags['gname'][n]]
        d = graph_data[n]
        for fld in fit_keys:
            df = d[fld]
            row.append(Fmt(df, fmt_keys[fld], strlen, sigfig))

        stats_data.append(row)
    # first row and col could be different sizes
    colW = [1.5 * cm]
    colW.extend([1.75 * cm] * (len(fit_keys))) # +2))
    rowH = [0.5 * cm]
    rowH.extend([0.5 * cm] * (numGraphs + 1))
    tabW = sum(colW)
    tabH = sum(rowH)
    yTab = yTop - tabH
    # use tabStyleBase, with additions
    tabStyle = tabStyleBase()
    for n in range(2):
        eCol = 5 * (n + 1) # 4 *
        sCol = eCol - 4 # 3
        tabStyle.append(('SPAN', (sCol, 0), (eCol, 0)))
        for i in range(4): # 3
            tabStyle.append(('LINEBEFORE', (eCol-i, 0), (eCol-i, -1), 0.5, colors.grey))

    t=Table(stats_data, colWidths=colW, rowHeights=rowH, style=tabStyle)

```

```

t.wrapOn(canv, tabW, tabH)
t.drawOn(canv, xLeft, yTab)
return yTab

def CrossStatsTab(canv, numGraphs, xLeft, yTop, graph_tags, graph_data, test_keys):
    # the f-tests comparing fits for different graphs
    stats_data = [[graph_tags['graph']]
    for n in range(0, numGraphs):
        stats_data[0].extend([graph_tags['gname'][n], ""])
    stats_data.append([""])
    stats_data[1].extend([graph_tags['test'], graph_tags['val']] * numGraphs)
    # should have well behaved numbers but sometimes not
    # a bit of fancy formatting...
    sigfig = 5
    strlen = 7
    for n1 in range(numGraphs):
        row = [graph_tags['gname'][n1]]
        for n2 in range(numGraphs):
            if n1 == n2:
                row.extend(["", ""])
            else:
                d = graph_data[n1][n2]
                dt = d[test_keys['test']]
                dv = d[test_keys['val']]
                row.append(Fmt(dt, "#", strlen, sigfig))
                row.append(Fmt(dv, "#", strlen, sigfig))

        stats_data.append(row)

    # first row and col could be different sizes
    colW = [1.5 * cm]
    colW.extend([1.5 * cm] * numGraphs * 2)
    rowH = [0.5 * cm]
    rowH.extend([0.5 * cm] * (numGraphs + 1))
    tabW = sum(colW)
    tabH = sum(rowH)
    yTab = yTop - tabH
    # use tabStyleBase, with additions
    tabStyle = tabStyleBase()
    for n in range(numGraphs):
        eCol = 2 * (n + 1)
        sCol = eCol - 1
        tabStyle.append(('SPAN', (sCol, 0), (eCol, 0)))
        tabStyle.append(('LINEBEFORE', (eCol, 0), (eCol, -1), 0.5, colors.grey))
    t = Table(stats_data, colWidths=colW, rowHeights=rowH, style=tabStyle)
    t.wrapOn(canv, tabW, tabH)
    t.drawOn(canv, xLeft, yTab)
    return yTab

def tabStyleBase():
    # reportlab greys (dec|hex): [NB: 'darkgrey' is lighter than 'grey']
    # dimgrey(105|69); grey(128|80); darkgrey(169|A9); lightgrey(211|D3)
    # N.B. these aren't the same as the R greys

    tabStyleB = [
        ('BOX', (0, 0), (-1, -1), 1.5, colors.black),
        ('GRID', (0, 0), (-1, -1), 0.5, colors.black),
        ('ALIGN', (0, 0), (-1, -1), 'RIGHT'),
        ('LINEBELOW', (0, 1), (-1, 1), 1, colors.dimgrey),
        ('ALIGN', (0, 0), (-1, 0), 'CENTRE'),
        ('ALIGN', (0, 1), (-1, 1), 'CENTRE'),
        ('LINEAFTER', (0, 0), (0, -1), 1, colors.dimgrey),
        ('ALIGN', (0, 0), (0, -1), 'CENTRE'),
        ('SPAN', (0, 0), (0, 1)),
    ]
    return tabStyleB

```

TabsToXL.py

```
# -----
# TabsToXL.py
# -----

import sys, string, os, arcgisscripting #, win32com.client
from xlwt import Workbook
gp = arcgisscripting.create(9.3)

def main(*argv):
    strScrName = "TabsToXL"
    strCodeSec = "getting parameters"
    strExtraInfo = ""
    try:

        # Script arguments...
        strCmdArgs = ""

        iParam=0
        in_TabList = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_TabList == '#': in_TabList = ""
        strCmdArgs = strCmdArgs + " " + in_TabList
        iParam=iParam+1
        in_XLFile = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_XLFile == '#': in_XLFile = ""
        strCmdArgs = strCmdArgs + " " + in_XLFile
        iParam=iParam+1

        in_KTO = argv[iParam] # gp.GetParameterAsText(iParam)
        if in_KTO == '#': in_KTO = ""
        strCmdArgs = strCmdArgs + " " + in_KTO
        iParam=iParam+1

        blnKTO = (in_KTO.lower() == "true")

        strExcel = in_XLFile

        lstTabs = in_TabList.split(";")
        strTmpView = "tmpXL"

        topRow = 0
        leftCol = 0

        xlBook = Workbook()

        for strTab in lstTabs:
            strTabName = gp.Describe(strTab).BaseName

            gp.MakeTableView_management(strTab, strTmpView)

            fields = gp.ListFields(strTmpView)

            fieldNames = []
            for field in fields:
                if (field.type <> "Geometry" and field.type <> "BLOB"):
                    fieldNames.append(field.name)

            rowNum = topRow
            columnNum = leftCol

            xlSheet = xlBook.add_sheet(strTabName)

            for fieldName in fieldNames:
                xlSheet.write(rowNum, columnNum, fieldName)
                columnNum = columnNum + 1

            cur = gp.SearchCursor(strTmpView)
            row = cur.Next()
            while row:
                rowNum = rowNum + 1
                columnNum = leftCol # 1
                for fieldName in fieldNames:
                    xlSheet.write(rowNum, columnNum, row.GetValue(fieldName))
                    columnNum = columnNum + 1

                row = cur.Next()

            del cur

        try:
            xlBook.save(strExcel)
        except:
            gp.AddMessage("Exception trying to save file!\n%s" % strExcel)
            raise
```



```
### Done ###
gp.AddMessage("Done.")
strCodeSec = "End!"
strExtraInfo = ""

if blnKTO:
    strCodeSec = "exception to keep tool open!"
    exit() # keep tool open in case works and gets to here!

except:
    gp.AddMessage("Got exception in %s, section = %s %s" % (strScrName, strCodeSec, strExtraInfo))
    raise
else:
    return 0 # exit errorlessly

if __name__ == '__main__':
    main(*sys.argv[1:])
```

ExtractData2CSV.py

```
# -----
# ExtractData2CSV.py
# -----

# script doesn't use parameters
# - set manually for data required

import arcgisscripting

import LatitudeFunctions as LF
import LatitudeDataNames as LDN

# Create the Geoprocessor object
gp = arcgisscripting.create(9.3)

# True - write file
# False - for debugging rest
blnWrite = True # False

lstStdPop = ["euro"]
strStdPop = LF.parseStdPop(lstStdPop[0])

# resident criteria
lstResPeriod = ["all"]
strRPer, strRPeriod, RPerYears, RPerMonths = LF.parsePeriod(lstResPeriod[0])

# factors
lstFacType = ["birth_month"]
strFacType = LF.parseFacType(lstFacType[0])

lstFactor = ["1", "2", "3", "4", "5", "6", "7", "8", "9", "j", "k", "l"]
lutFac = dict( [m, m] for m in lstFactor)
# map oct-dec codes to month numbers
lutFac["j"] = "10"
lutFac["k"] = "11"
lutFac["l"] = "12"

# prefilters
lstPreFs = ["res_all_type23", "res_all_type4"]

# life periods
lstLifePeriod = ["0a4", "4o0", "prv"]
#lstLifePeriod = ["0a4"]

# genders
lstGUFac = ["aa", "fa", "ma"]

# lat bin sizes
lstLatBins = ["4.0", "3.0", "2.0"]

yr = { '0a4': "1956", 'prv': "2006", '4o0': "2006" }

strBaseDir = "D:/GIS/Data"
strScratchWDS = "D:/Workspace/Scratch.mdb"

strLPWDS = strBaseDir + "/" + "LatPops.mdb"

strDataCSV = strBaseDir + "/" + strFacType + "_data.csv"

# fields

strFld_lc = LDN.FieldLatC()
strFld_bin = "bin_size"
# factors
strFld_gen = LDN.FieldGender()
strFld_prf = "pre_factor"
strFld_fty = "factor_type"
strFld_fac = LDN.FieldFactor()
strFld_rpr = LDN.FieldResPer()
strFld_lpr = LDN.FieldLifePer()
strFld_cyr = LDN.FieldCYear()
strFld_std = LDN.FieldStdPop()
strFld_smp = "sample_size"
strFld_pop = "lat_pop"
strFld_num = "lat_num"

strCodeSec = ""
strExtraInfo = ""
try:

    if blnWrite:
        # csv version
        strCodeSec = "writing header"
        strExtraInfo = "; opening output %s" % strDataCSV
        f=open(strDataCSV, 'w')
        strExtraInfo = "; writing header line"
        f.write(','.join([strFld_prf, strFld_fty, strFld_fac, strFld_rpr, strFld_lpr,
```

```

        strFld_gen, strFld_cyr, strFld_std, strFld_bin, strFld_smp,
        strFld_lc, strFld_pop, strFld_num ] ))
f.write("\n")

for strPref in lstPrefFs:
    strCodeSec = "strPref loop"
    strExtraInfo = "; strPref = %s" % strPref
    strEDWDS = "/" . join([strBaseDir, strPref, strFacType, "tmp_ED.mdb"])

for lp in lstLifePeriod:
    strCodeSec = "lp loop"
    strExtraInfo = "; lp = %s" % lp
    strLPer, strLPeriod, LPerYears, LPerMonths = LF.parsePeriod(lp)
    strCYear = LF.parseCYear(yr[strLPer])

    for fac in lstFactor:
        strCodeSec = "fac loop"
        strExtraInfo = "; fac = %s" % fac
        strFactor = LF.parseFactor(fac)

    for strLBin in lstLatBins:
        strCodeSec = "strLBin loop"
        strExtraInfo = "; strLBin = %s" % strLBin
        bfac = 1.0 / float(strLBin)
        strLatBin = LF.parseLat(strLBin)

# need different loops:
# lat pops - one table per gufac
# lat nums - one table for all gufacs
# (and can't have two search cursors at same time)

# lat pops:
Ylp = dict( [gu, dict()] for gu in lstGUFac)
for gu in lstGUFac:
    strCodeSec = "gu loop"
    strExtraInfo = "; gu = %s" % gu
    strGUFac = LF.parseFactor(gu)
    strFld_lp = LDN.AllFieldsLatPop(strGUFac, strCYear)

    strCodeSec = "setting tables"
    strExtraInfo = "strPLatsTab"
    # input tables of census population per latitude bin
    strPLatsTabName = LDN.KeepLatPop(strCYear, strLatBin, strGUFac)
    strPLatsTab = strLPWDS + "/" + strPLatsTabName

# NA handling broken in 2.0 versions of RPy2
# replaced by NA_Real etc in 2.1 versions
# so instead of setting to NA equiv just don't add to lists
# however, must use matching LatC list for data
# whether different files or Y fields in same file
# note: with lat pops split across files, not all files have all gu's
# also: skipping NA/0 also good for logit!

    strExtraInfo = "search cursor %s" % (strPLatsTab)
    get_rows = gp.SearchCursor(strPLatsTab, "", "", "", strFld_lc + " A")
    get_row = get_rows.Next()

    if not get_row:
        gp.AddMessage("Cannot get data from %s" % strPLatsTab)
        strExtraInfo = "keeping tool open"
        exit() # keep tool open

    while get_row:

        raw_val = 0
        Xlat = get_row.GetValue(strFld_lc)
        for fY in strFld_lp:
            strExtraInfo = "file %s; fY (%s) %s"%(strPLatsTab, type(fY), fY)
            raw_val = get_row.GetValue(fY)

            if raw_val != 0:
                strExtraInfo = "file %s; raw_val %s * %s" % (strPLatsTab,
                    repr(raw_val), repr(fac))

                val = raw_val * bfac
                strExtraInfo = "file %s; fieldX %s"%(strPLatsTab, strFld_lc)
                strExtraInfo = "file %s; Ylp %s" % (strPLatsTab, repr(Ylp))
                if not Ylp[gu].has_key(Xlat):
                    Ylp[gu][Xlat] = 0.0
                Ylp[gu][Xlat] += val

        strExtraInfo = "file %s next" % (strPLatsTab)
        get_row = get_rows.Next()

    # done with the cursor
    del get_row
    del get_rows

# lat nums:
    strCodeSec = "setting tables"
    strExtraInfo = "strBLatsTab"
    # input tables of numbers of cases per latitude bin

```

```

strBLatsTabName = LDN.KeepLateNums(strRPer, strLPer, strLatBin, strFactor)
strBLatsTab = strEDWDS + "/" + strBLatsTabName

strExtraInfo = "search cursor %s" % (strBLatsTab)
get_rows = gp.SearchCursor(strBLatsTab, "", "", "", strFld_lc + " A")
get_row = get_rows.Next()

if not get_row:
    gp.AddMessage("Cannot get data from %s" % strBLatsTab)
    strExtraInfo = "keeping tool open"
    exit() # keep tool open

Ybn = dict( [gu, dict()] for gu in lstGUFac)
Ysm = dict( [gu, 0.0] for gu in lstGUFac)

while get_row:
    Xlat = get_row.GetValue(strFld_lc)

    for gu in lstGUFac:
        strGUFac = LF.parseFactor(gu)
        strFld_lno = LDN.FieldLatNum(strGUFac)

        fY = strFld_lno
        strExtraInfo = "file %s; fY (%s) %s" % (strBLatsTab, type(fY), fY)
        raw_val = get_row.GetValue(fY)

        if raw_val != 0:
            strExtraInfo = "file %s; raw_val %s * %s" % (strBLatsTab,
                                                         repr(raw_val), repr(fac))

            val = raw_val * bfac
            strExtraInfo = "file %s; Ybn %s" % (strBLatsTab, repr(Ybn))
            Ybn[gu][Xlat] = val
            Ysm[gu] += raw_val

        strExtraInfo = "file %s next" % (strBLatsTab)
        get_row = get_rows.Next()

# done with the cursor
del get_row
del get_rows

# should have key combinations the same for both Ylp and Ybn
XlatKeys = set()
for k in Ylp.keys(): XlatKeys = XlatKeys.union(Ylp[k].keys())
for k in Ybn.keys(): XlatKeys = XlatKeys.union(Ybn[k].keys())

for gu in lstGUFac:
    strGUFac = LF.parseFactor(gu)
    strGen = strGUFac[0]

    for Xlat in sorted(XlatKeys, reverse=True):
        X = "%g" % Xlat
        if Ylp[gu].has_key(Xlat): Y1 = "%g" % Ylp[gu][Xlat]
        else: Y1 = ""
        if Ybn[gu].has_key(Xlat): Y2 = "%g" % Ybn[gu][Xlat]
        else: Y2 = ""
        YS = "%g" % Ysm[gu]

        if blnWrite:
            f.write(', '.join([strPref, strFacType, lutFac[strFactor],
                              strRPer, strLPer, strGen, strCYear,
                              strStdPop, strLBin, YS, X, Y1, Y2 ]))
            f.write("\n")

    if blnWrite:
        f.close()

### Done ###
gp.AddMessage("Done.")
strCodeSec = "End!"
strExtraInfo = ""

except:
    gp.AddMessage("Got exception in Extractdata2CSV, section = %s %s"%(strCodeSec, strExtraInfo))
    raise
else:
    return 0 # exit errorlessly

```

Appendix D Least Squares Regression

This appendix will detail linear regression and the use of ANOVA to compare regression fits in section D.1, within the context of this study. Log regression will be described in section D.2 followed by the associated logit or logistic regression. Two excellent books for statistical analysis techniques in epidemiological studies are "Epidemiology" by Rothman (2002) and "An introduction to medical statistics", Bland (2000).

D.1 Linear Least Squares Regression

A standard linear least squares regression fit with independent and dependent variables x and y gives the expected rates \hat{y} as

$$\hat{y} = \beta_0 + \beta_1 x \tag{D.1}$$

where the fit variables, β_0 and β_1 , represent the intercept and gradient, respectively.

This equation is the *model*. The residuals, $(\hat{y} - y)$, are also considered which give the variance

$$s^2 = \frac{1}{n-2} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \tag{D.2}$$

and the correlation coefficient in terms of the means \bar{x} and \bar{y} is

$$r^2 = \frac{(\sum (x - \bar{x})(y - \bar{y}))^2}{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2} \tag{D.3}$$

The correlation coefficient is often regarded as a measure of the goodness of fit of a regression line—it represents the amount of variation in the data which can be attributed to the model underlying the regression line.

As well as looking at the gradients to see whether or not there is an effect of latitude (x) on the prevalence of MS cases (y), where $\beta_1 \neq 0$ indicates that there is an effect) the fits obtained for different graphs, or data sets, (either from different samples, or the same samples at different times) also need to be compared in some way. However two individual fits obtained for two given graphs cannot easily be compared. Instead, a further model is considered which incorporates both sets of data at the same time. (For simplicity only pair-wise comparisons are considered.) To do this an indicator or

dummy variable, z , is used which takes only two values: 1 when the data is from graph (data set) 1 and 0 when the data is from graph (data set) 2. The model then becomes:

$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 xz \quad \text{D.4}$$

which reduces to the following pair of equations when z is set to 1 and 0 respectively:

$$\begin{aligned} \hat{y}_{z=1} &= (\beta_0 + \beta_2) + (\beta_1 + \beta_3)x \\ \hat{y}_{z=0} &= \beta_0 + \beta_1 x \end{aligned} \quad \text{D.5}$$

Although this appears to be much the same as the two individual fits for the two graphs, it allows a direct comparison of the two graphs by looking at β_2 and β_3 . There are several options for the null hypothesis: the use of normalised, or age-standardised, data implies that if either of the intercepts or the gradients are the same then the lines should be coincident. That is, if either $\beta_2 = 0$ or $\beta_3 = 0$ then it should also be the case that $\beta_2 = \beta_3 = 0$. Any of these conditions could be checked for and for well behaved data the choice should not make much of a difference. However, the data in this study is often not well behaved so the slightly less strict test of parallelism, $H_0 : \beta_3 = 0$, is therefore chosen rather than the stricter test of coincidence, $H_0 : \beta_2 = \beta_3 = 0$. Note that this model assumes that the variances are equal within the two data sets.

To test H_0 the ANOVA tables for the regression model are used, and the values for the various tests can be read directly from the ANOVA tables as shown in Table D.1. A detailed discussion of ANOVA tables is beyond the scope of this dissertation, however, a few definitions and equations are given here for reference. They are given without further explanation for the linear regression model with k explanatory variables:

RSS_k is the residual variation (sums of squares):

$$RSS_k = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \text{D.6}$$

S_{yy} is the total variation (sums of squares):

$$S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad \text{D.7}$$

SSS_j is the sequential sum of squares, the variation (sum of squares) due to variable j in the model and is given by:

$$SSS_j = RSS_{j-1} - RSS_j \tag{D.8}$$

For the $H_0 : \beta_3 = 0$ model, with $k = 3$, the F -test statistic for H_0 reduces to

$$F_{test} = \frac{SSS_3}{RSS_3/(n-4)} \tag{D.9}$$

which is then compared to an $F(1, n-4)$ distribution to get the p -value. If the stricter $H_0 : \beta_2 = \beta_3 = 0$ is chosen then the F -test statistic for H_0 is

$$F_{test} = \frac{(SSS_2 + SSS_3)/2}{RSS_3/(n-4)} = \frac{(S_{yy} - RSS_3 - SSS_1)/2}{RSS_3/(n-4)} \tag{D.10}$$

which is compared to an $F(2, n-4)$ distribution to get the p -value.

It should also be noted here that comparisons between linear fits for different data sets can really only be carried out successfully if the populations being studied are of a similar size. This is discussed in more detail in section 4.3.3.1. The next section in this Appendix discusses the alternative log and logistic regression techniques.

Source	d.f.	Sum of Squares	Mean Square	F -value	p -value
Model	k	$S_{yy} - RSS_k$	$(S_{yy} - RSS_k)/k$	$\frac{(S_{yy} - RSS_k)/k}{RSS_k/(n-k-1)}$	
Error	$n-k-1$	RSS_k	$RSS_k/(n-k-1)$		
Total	$n-1$	S_{yy}			
Source	d.f.	Seq. sum of Squares			
β_1	1	$S_{yy} - RSS_1$			
β_2	1	$RSS_1 - RSS_2$			
\vdots	\vdots	\vdots			
β_j	1	$RSS_{j-1} - RSS_j$			
\vdots	\vdots	\vdots			
β_{k-1}	1	$RSS_{k-2} - RSS_{k-1}$			
β_k	1	$RSS_{k-1} - RSS_k$			

Table D.1 ANOVA table

A typical ANOVA table indicating the variables which are output. d.f. is the degrees of freedom; other variables are as described or defined in the text.

D.2 Log and Logistic Least Squares Regression

As mentioned at the beginning of section 4.3.3 on regression techniques, linear regression is often used to look at variations in the rates of prevalence with latitude, despite this technique being less than ideal for the rate data being studied. This section will look at the techniques of log and logistic regression which are much more suited to looking at rate data.

If there were just two measures, prevalence in the north and prevalence in the south, for some definition of north and south, then for a given sample either a case comes from the north or it comes from the south. As such, it would be possible to analyse the risk odds ratio for the data. However, there are two problems with that approach in this study. First, there is no control data, which is required for looking at risk odds. This will be considered later. In addition, this study is interested in variation of prevalence with latitude which is a continuous variable rather than dichotomous. An approach which does allow for a continuous variable is logistic regression, which can be used instead of linear regression. As well as being directly related to the risk odds ratio, it has the added advantage of removing one major disadvantage which is inherent in linear regression—whereas linear regression can predict values which are negative or very large, predictions from logistic regression are constrained to the range [0,1].

The methodology involved for logistic regression will be described later in this section; first it is useful to look at the simpler log regression. Note that the log function used in the regression can be to any base without loss of generality—here the natural log function $\ln()$ is used along with the inverse function $\exp()$; if the base 10 log function $\log()$ is used then the inverse function is the power of 10.

Log least squares regression is very similar to linear least squares regression. For convenience, let the dependent variable be s and $y = \ln(s)$. The regression model, then becomes:

$$\hat{y} = \ln(\hat{s}) = \beta_0 + \beta_1 x \tag{D.11}$$

or:

$$\hat{s} = \exp(\beta_0 + \beta_1 x) \tag{D.12}$$

and testing of the log regression model of equation D.11 is carried out in exactly the same way as with the linear regression model described in section D.1.

Consider the ratio of \hat{s} for two values of the independent variable, x_1 and x_2 :

$$\frac{\hat{s}_1}{\hat{s}_2} = \frac{\exp(\beta_0 + \beta_1 x_1)}{\exp(\beta_0 + \beta_1 x_2)} = \exp(\beta_1(x_1 - x_2))$$

D.13

For a unit change in x , this reduces to $\exp(\beta_1)$. Thus $\exp(\beta_1)$ is the expected rate *ratio* of predicted values for a unit change in the independent variable. This compares with the linear regression coefficient β_1 which gives the expected rate *change* in predicted value for a unit change in the independent variable. As mentioned briefly at the end of the last section, comparisons of fits from linear regression models should be based on the same underlying populations. This is no longer such an issue with log regression as the fit coefficient represents a ratio rather than an absolute change. This is discussed further in section, 4.3.3.1.

One of the measures of interest within this study is the ratio F:M of female to male prevalence rates. Assuming that the log regression functional form is suitable for the data, then the ratio of rates for a given latitude will be similar to the ratio looked at in D.13, though now there is a single value of x and two different fits F and M for female and male, respectively. This gives:

$$\frac{\hat{s}_F}{\hat{s}_M} = \frac{\exp(\beta_{0F} + \beta_{1F}x)}{\exp(\beta_{0M} + \beta_{1M}x)} = B_{FM} \exp((\beta_{1F} - \beta_{1M})x)$$

D.14

where $B_{FM} = \exp(\beta_{0F} - \beta_{0M})$. For two fits to have the same F:M ratio for different x :

$$B_{FM} \exp((\beta_{1F} - \beta_{1M})x_1) = B_{FM} \exp((\beta_{1F} - \beta_{1M})x_2)$$

D.15

which implies:

$$\exp(\delta\beta_1 \delta x) = 1$$

D.16

where $\delta\beta_1 = (\beta_{1F} - \beta_{1M})$ and $\delta x = (x_1 - x_2)$, which can only hold for two different x if $\beta_{1F} = \beta_{1M}$. That is, the F:M ratio will vary with latitude unless the female and male rate ratios are equal.

Whereas linear least squares regression fits a straight line to the data, it can be seen from equations D.11 and D.12 that log regression is equivalent to fitting an exponential curve to the data. Since the exponential function cannot go negative, at least one of the limitations on the use of linear regression for rate data has been removed with log regression. The other main restriction of linear and log regressions, that of having predicted rates greater than 1, can be removed by extending the log regression model to one of logistic regression.

Log regression is an extension of linear regression with dependent variable s being modelled as $y = \ln(s)$. Logistic regression is similar and makes use of the $\text{logit}()$ function instead of the $\ln()$ function. If p is a proportion in the range $[0,1]$ then the $\text{logit}()$ function is defined as:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) \tag{D.17}$$

The logistic regression fit then takes the form

$$\hat{y} = \text{logit}(\hat{p}) = \beta_0 + \beta_1 x \tag{D.18}$$

from which the expected or predicted proportions \hat{p} are given by:

$$\hat{p} = \frac{\exp(\hat{y})}{1 + \exp(\hat{y})} \tag{D.19}$$

The predicted \hat{y} values can still be negative or very large, however, the $\text{logit}()$ function transforms the full range of \hat{y} values to the range $[0,1]$, which is suitable for rate data. In addition, if the regression equation D.18 is transformed by taking the exponential of each side, then an equivalent of the risk odds is obtained:

$$\frac{\hat{p}}{1-\hat{p}} = \exp(\beta_0 + \beta_1 x) \tag{D.20}$$

The risk odds ratio for two values of the independent variable, x_1 and x_2 , is then

$$\frac{\hat{p}_1(1-\hat{p}_2)}{\hat{p}_2(1-\hat{p}_1)} = \frac{\exp(\beta_0 + \beta_1 x_1)}{\exp(\beta_0 + \beta_1 x_2)} = \exp(\beta_1(x_1 - x_2)) \tag{D.21}$$

which, for a unit change in x , reduces to $\exp(\beta_1)$, as it did in equation D.13 for the log regression. So for logistic regression, $\exp(\beta_1)$ can be considered as the expected risk odds ratio for a unit change in the independent variable and is discussed in more detail in 4.3.3.1.

For both log and logistic regression, comparisons between data sets can be made in the same manner as for linear regression, with the introduction of a dummy variable. With these, however, the model can only be used to check for parallelism rather than for full coincidence, so $H_0 : \beta_3 = 0$ is the only option. The F -test statistic for H_0 is the same as for the linear regression test and is given by equation D.6. F_{test} is again compared to an $F(1, n - 4)$ distribution to get the p -value.

It is debatable which of log or logistic regression should be used as an alternative to linear regression here. Logistic regression is suitable for rate data, yet it is based on proportions, which require control data not present in this study. Log regression does not need control data but is not entirely suitable for rate data. It is interesting to note, however, the similarity of the $\ln()$ and $\text{logit}()$ functions for the data being considering here. The difference, δ , between the two functions for a given value of p is:

$$\delta = \text{logit}(p) - \ln(p) = \ln\left(\frac{p}{1-p}\right) - \ln(p) = -\ln(1-p) \quad \text{D.22}$$

The Taylor series for $\ln(1+x)$ around $x_0 = 0$ (that is, for small x) is:

$$\ln(1+x) = \sum_{i=1}^{\infty} \frac{(-1)^{i+1}}{i} x^i = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \dots \quad \text{D.23}$$

which gives:

$$\delta = -\ln(1-p) = p + \frac{1}{2}p^2 + \frac{1}{3}p^3 + \frac{1}{4}p^4 + \dots \quad \text{D.24}$$

Although the series D.23 and D.24 are very slowly convergent for most of the range which x or p can take $(-1,1)$, the highest prevalence rates in this study are no more than a few hundred per 100,000. Consider a maximum p_{\max} value of, say, 0.005. The first two terms in the series in equation D.24 will be 0.005 and 1.25e-5, so the following approximation can then be made to four decimal places:

$$\delta = \text{logit}(p) - \ln(p) \approx p \quad p \leq p_{\max}$$

D.25

As confirmation of this result, the values for $p = 0.005$ to five significant figures are:

$$\text{logit}(0.005) = -5.2933$$

$$\ln(0.005) = -5.2983$$

$$\delta = 5.0125 \times 10^{-3}$$

That is, the $\ln()$ and $\text{logit}()$ functions differ from each other by about 0.09% for $p = 0.005$. Note that as p gets smaller and closer to 0, from equation D.24 δ also gets smaller. In addition, $\ln(p)$ and $\text{logit}(p)$ both get larger negative values so the percentage difference will decrease even further. Therefore, the previous statement can be generalised to state that the $\ln()$ and $\text{logit}()$ functions differ from each other by less than 0.1% for p in $[0, 0.005]$.

In conclusion, for the data being studied here, either log or logistic regression are likely to give the same results within the uncertainties of the data and the other methods involved in this study. Given the lack of control data in this study log regression will be used and $\exp(\beta_1)$ will be taken as the rate ratio for a unit change in the independent variable.

Appendix E Migration Models

The classic 'gravity' model (Zipf 1946, 1949) has evolved somewhat over the years (Cohen et al. 2008), yet it is simple enough to adapt to the analysis required here.

In the following simplified analysis, it is assumed that there are no births or deaths and that there is no migration into or out of the regions from or to areas outside.

Consider two regions A and B. For a given time period there will be a change in the populations of A and B as follows:

$$\begin{aligned} P'_A &= P_A + P_{A \leftarrow B} - P_{B \leftarrow A} \\ P'_B &= P_B + P_{B \leftarrow A} - P_{A \leftarrow B} \end{aligned} \tag{E.1}$$

where P_A , P_B , P'_A and P'_B are the original and the new populations, respectively, and $P_{A \leftarrow B}$ and $P_{B \leftarrow A}$ are the numbers migrating to A from B and to B from A, respectively.

In matrix form this can be written:

$$\begin{bmatrix} P'_A \\ P'_B \end{bmatrix} = \begin{bmatrix} \frac{(P_A - P_{B \leftarrow A})}{P_A} & \frac{P_{A \leftarrow B}}{P_B} \\ \frac{P_{B \leftarrow A}}{P_A} & \frac{(P_B - P_{A \leftarrow B})}{P_B} \end{bmatrix} \begin{bmatrix} P_A \\ P_B \end{bmatrix} \tag{E.2}$$

The elements of the matrix are thus seen to be the proportions of the population of one region migrating to another region, which, in the case of the diagonal elements, is the proportion staying in the region and not migrating out. Equation E.2 can be generalised to give the matrix equation:

$$P' = MP \tag{E.3}$$

with population and migration matrices, P , P' and M , as follows:

$$\begin{aligned} P &= \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{bmatrix}; \quad P' = \begin{bmatrix} p'_1 \\ p'_2 \\ \vdots \\ p'_N \end{bmatrix} \\ M &= \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1N} \\ m_{21} & m_{22} & \cdots & m_{2N} \\ \vdots & \vdots & & \vdots \\ m_{N1} & m_{N2} & \cdots & m_{NN} \end{bmatrix} \end{aligned} \tag{E.4}$$

where p_j and p'_j are the original and new populations for region j ; m_{jk} ($k \neq j$) is the proportion moving from region k to region j ; and m_{jj} is the proportion staying in region j , given by:

$$m_{jj} = 1 - \sum_{i \neq j} m_{ij} \quad \text{E.5}$$

The migration matrix M is then applied to the population matrix P for as many time steps as required to give a final population matrix P' . It is important to note that the elements m_{jk} may not be constant proportions. The classic 'gravity' model takes m_{jk} of the form:

$$m_{jk} = \frac{\alpha p_j^\beta}{d_{jk}} \quad (k \neq j) \quad \text{E.6}$$

where d_{jk} is the distance between regions, and $\beta > 0$, so that people are more likely to migrate shorter distances and from smaller to larger centres of population. In this case the matrix M must be recalculated at each time step based on the new p_j values.

The census populations discussed in Chapter 4 show an almost doubling of the population from 1956 to 2006. However, the change is not uniform with the greatest increase in population being in the Auckland area. Although much of this overall increase is due to both higher birth rates than death rates, and migration from outside of New Zealand (all factors which are not considered in the simple model described here), the change in the distribution of the population can be modelled as follows. Note that this model is not an attempt to describe the details of internal migration, but rather is a mechanism to predict how a sub-population might be redistributed. As such, no physical meaning should be associated with the parameters used in the model.

Let the initial population matrix P take the shape of the 1956 census population distribution shown in Figure 4.5(d) with seven regions corresponding to the seven 2° latitude slices shown in Figure 4.5. Also, let the migration matrix parameters in equation E.6 be $\alpha = 1$, $\beta = 2$ and $d_{jk} = (|j - k| + 0.5)$, and recalculate and apply M to P a total of seven times. (Note that each time step is not necessarily equivalent to a

year—the number of time steps, and the values for α , β and d_{jk} , were chosen simply to provide a final population matrix P' similar in shape to the 2006 population distribution. Several other combinations of values were also used, however the combination reported here gave the closest match.)

P and the final P' matrix are then given by:

$$P = \begin{bmatrix} 34271 \\ 294420 \\ 182879 \\ 212814 \\ 136359 \\ 102589 \\ 36668 \end{bmatrix} \times 10^{-6} ; \quad P' = \begin{bmatrix} 16956 \\ 410110 \\ 138227 \\ 219546 \\ 107727 \\ 080652 \\ 26782 \end{bmatrix} \times 10^{-6}$$

E.7

These distributions are shown in Figure 4.8(a), and it can be seen that the shape of the final distribution is similar to the population distribution for 2006 in Figure 4.5(d).

Combining the values of M for each time step into a single matrix M' gives:

$$M' = \begin{bmatrix} 0.455820 & 0.002112 & 0.001432 & 0.00109 & 0.000883 & 0.000738 & 0.000632 \\ 0.370869 & 0.772589 & 0.361123 & 0.256423 & 0.201114 & 0.164139 & 0.137985 \\ 0.058720 & 0.082819 & 0.442148 & 0.082090 & 0.056852 & 0.043485 & 0.035123 \\ 0.081819 & 0.103352 & 0.146161 & 0.589952 & 0.151469 & 0.102579 & 0.077888 \\ 0.021519 & 0.026008 & 0.033323 & 0.049807 & 0.558042 & 0.051736 & 0.033719 \\ 0.010204 & 0.011923 & 0.014417 & 0.018920 & 0.029338 & 0.633619 & 0.030707 \\ 0.001048 & 0.001197 & 0.001397 & 0.001718 & 0.002301 & 0.003704 & 0.683947 \end{bmatrix}$$

E.8

If M' is applied to any sub-population P_{Sub} the resulting P'_{Sub} should indicate how P_{Sub} would be re-distributed if the only effects in force were inter-regional migration effects.

Consider the following distribution P_{Sub} with rates per 100,000 given by R_{Sub} :

$$P_{Sub} = \begin{bmatrix} 14 \\ 149 \\ 114 \\ 160 \\ 124 \\ 112 \\ 49 \end{bmatrix} \times 10^{-6} ; \quad R_{Sub} = \begin{bmatrix} 40.9 \\ 50.6 \\ 62.3 \\ 75.2 \\ 90.9 \\ 109 \\ 134 \end{bmatrix}$$

E.9

Applying M' gives new sub-populations and rates, to three significant figures:

$$P'_{Sub} = \begin{bmatrix} 7.26 \\ 253 \\ 90.3 \\ 162 \\ 92.6 \\ 82.7 \\ 34.8 \end{bmatrix} \times 10^{-6} ; \quad R'_{Sub} = \begin{bmatrix} 42.8 \\ 61.6 \\ 65.4 \\ 73.6 \\ 85.9 \\ 103 \\ 130 \end{bmatrix}$$

E.10

R_{Sub} and R'_{Sub} are shown in Figure 4.8(b) along with log fits to the data.

Appendix F Solar Insolation

There are many web sites which provide simplified equations for solar radiation exposure; most of these are directed at users of photovoltaic solar arrays. One such web site is pveducation.org (pveducation n.d.) from which the following is based.

Note that no attempt is made here to include any atmospheric effects; the exposures described here would also vary with altitude, cloud cover, ozone and various other atmospheric conditions described in section 6.2. In addition, the absorption of solar radiation by the atmosphere depends on the solar elevation (and hence time of year and day) and is wavelength dependent, with UV radiation being absorbed more than visible light. Although this means that no direct comparison should be made between these exposures for different times of the year, or times of the day for the daily cycles, they can be useful in illustrating how exposure varies with the angle of the surface.

An estimate of Solar Exposure for different latitudes and times of the year is defined by the following:

$$SolExp = Sol_{inc} \cos\left(\frac{\pi}{180}[\beta - lat + \delta]\right) \quad \text{F.1}$$

where

$$\delta = \delta_{max} \sin\left(\frac{2\pi}{365}[284 + d]\right) \quad \text{F.2}$$

and Sol_{inc} is the incident solar radiation, β is the angle of the surface being considered, d is the day of the year, lat is the latitude and δ_{max} is the maximum tilt angle of the earth's axis. Daily variations are given by

$$SolExp_D = Sol_{inc} \sin\left(\frac{\pi}{180}[\beta + \alpha]\right) \quad \text{F.3}$$

where

$$\alpha = \sin^{-1}(\sin(\delta) \sin(lat) + \cos(\delta) \cos(lat) \cos(Hr_{\alpha})) \quad \text{F.4}$$

and the hour angle, Hr_{α} (degrees), is given by

$$Hr_{\alpha} = 15 \left(h + \frac{T_{corr}}{60} - 12 \right) \quad \text{F.5}$$

The local time correction, T_{corr} , is given by

$$T_{corr} = 4 (lon - 15 TZ_{Hr}) + EoT \quad \text{F.6}$$

in terms of the longitude, lon , the time zone offset in hours, TZ_{Hr} and the 'Equation of Time' correction for the eccentricity of Earth's orbit given by

$$EoT = 9.87 \sin(2B) - 7.53 \cos(B) - 1.5 \sin(B) \quad \text{F.7}$$

with $B = (d - 81) / 365$.

The annual cycle of relative solar exposure is illustrated in Figure F.1 for 30°, 40° and 50° south for (a) a horizontal surface and (b) a vertical surface. The ratio of vertical to horizontal radiation exposure is given in Figure F.1(c). This illustrates that in winter, a vertical surface can have much greater exposure than a horizontal plane.

Daily exposure levels are shown in Figure F.2 for mid winter and mid summer at 30° and 50° south for (a) a horizontal surface and (b) a vertical surface. There is a similar pattern to the annual cycle of exposure but with cutoffs for sunrise and sunset.

[Figure F.1 and Figure F.2 are shown on the following pages]

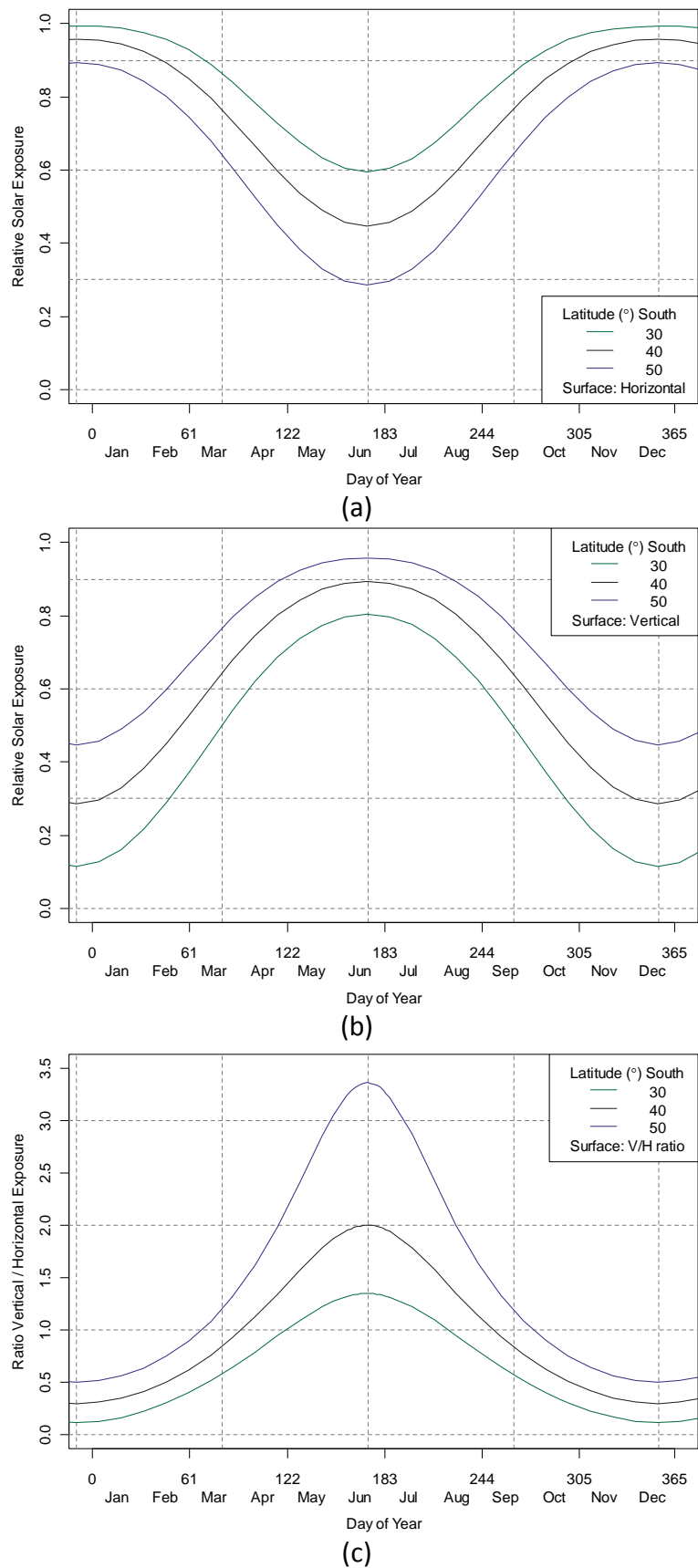
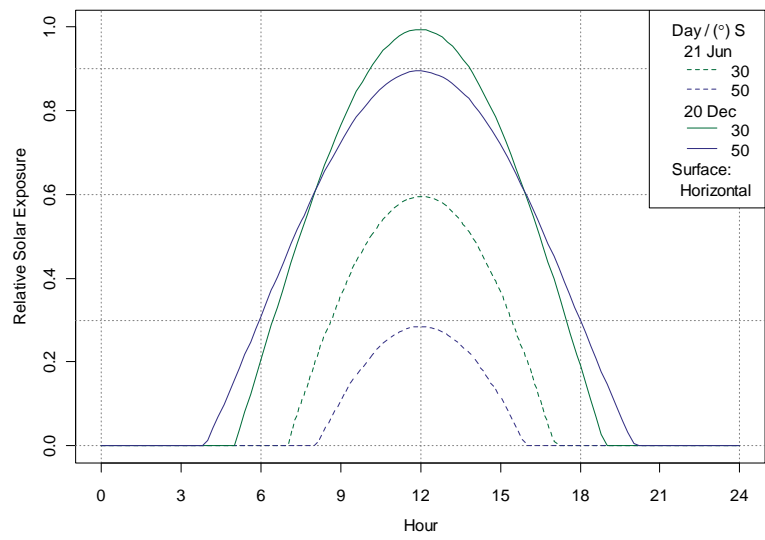
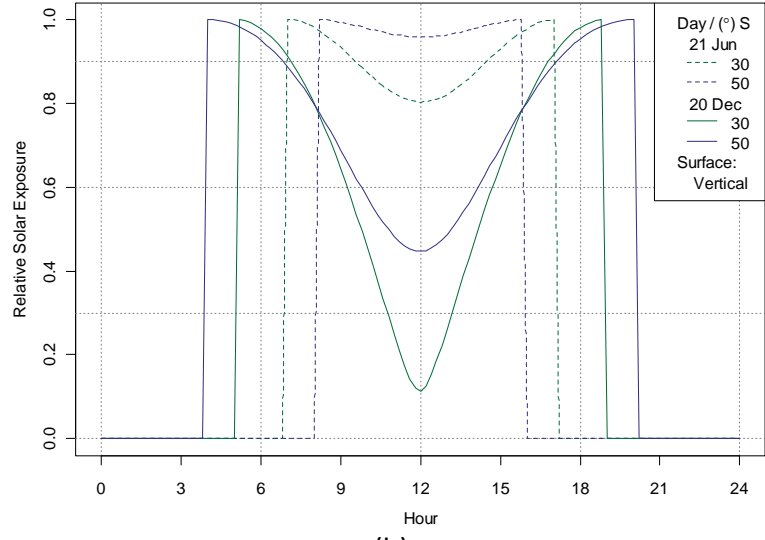


Figure F.1 Annual Variations in Relative Solar Radiation by Latitude

Annual variations in the relative solar exposure on a flat surface for latitudes of 30°, 40° and 50° south (from equation F.) for **(a)** a horizontal surface **(b)** a vertical surface and **(c)** the ratio of vertical to horizontal. The vertical grid lines are aligned with the solstice and equinox dates rather than the calendar months of the year.



(a)



(b)

Figure F.2 Daily Variations in Relative Solar Radiation by Latitude
 Daily variations in the relative solar exposure on a flat surface for latitudes of 30° and 50° south (from equation F.3) for mid winter and mid summer for **(a)** a horizontal surface and **(b)** a vertical surface.

References

Abbreviations used within the references:

AUAIHW	Australian Government, Australian Institute of Health and Welfare
AUDHA	Australian Government,, Department of Health and Ageing
CCA	Cancer Council Australia
CRUK	Cancer Research UK
DKMFLF	Ministeriet for Fødevarer, Landbrug Og Fiskeri
IMSGC	International Multiple Sclerosis Genetics Consortium
MSIF	Multiple Sclerosis International Federation
NZDoS	New Zealand Dept. of Statistics
NZMCH	New Zealand Ministry for Culture and Heritage Te Manatu Taonga
StatsNZ	Statistics New Zealand
WHO	World Health Organization
WTCCC	Wellcome Trust Case Control Consortium

Abad, P., Pérez, M., Castro, E., Alarcón, T., Santibáñez, R., and Díaz, F., 2010.

Prevalence of multiple sclerosis in Ecuador. *Neurología*, 25 (5), 309–313.

Adams, J. S., and Hewison, M., 2010. Update in vitamin D. *Journal of Clinical*

Endocrinology & Metabolism, 95 (2), 471–478.

Ahlgren, C., Odén, A., and Lycke, J., 2011. High nationwide prevalence of multiple sclerosis in Sweden. *Multiple Sclerosis Journal*, 17 (8), 901–908.

Ahmad, O. B., Boschi-Pinto, C., Lopez, A. D., Murray, C. J. L., Lozano, R., and Inoue, M., 2001. Age standardization of rates: A new WHO standard. *GPE Discussion Paper Series: No.31 EIP/GPE/EBD World Health Organization 2001*.

Akdis, M., Burgler, S., Crameri, R., Eiwegger, T., Fujita, H., Gomez, E., Klunker, S., Meyer, N., O'Mahony, L., Palomares, O., Rhyner, C., Quaked, N., Schaffartzik, A., Van De Veen, W., Zeller, S., Zimmermann, M., and Akdis, C. A., 2011. Interleukins, from 1 to 37, and interferon- γ : Receptors, functions, and roles in diseases. *Journal of Allergy and Clinical Immunology*, 127 (3), 701–721.e70.

Al-Hashel, J., Besterman, A. D., and Wolfson, C., 2008. The prevalence of multiple sclerosis in the Middle East. *Neuroepidemiology*, 31 (2), 129–137.

- Aladro, Y., Alemany, M. J., Pérez-Vieitez, M. C., Amela, R., Conde, M., Reyes, M. P., Álamo, F., and Angel-Moreno, A., 2005. Prevalence and incidence of multiple sclerosis in Las Palmas, Canary Islands, Spain. *Neuroepidemiology*, 24 (1–2), 70–75.
- Alshishtawy, M. M., 2012. Vitamin D deficiency: This clandestine endemic disease is veiled no more. *Sultan Qaboos University Medical Journal*, 12 (2), 140–152.
- Alstadhaug, K. B., Olavsen, J., and Salvesen, R., 2005. Forekomst av multippel sklerose i Nordland 1970–99 [Incidence and prevalence of multiple sclerosis in a Norwegian county, 1970–1999]. *Tidsskrift for Den Norske Lægeforening*, 125 (4), 431–433 (in Norwegian).
- Alter, M., and Harshe, M., 1975. Racial predilection in multiple sclerosis. *Journal of Neurology*, 210 (1), 1–20.
- Alter, M., Kahana, E., and Loewenson, R., 1978. Migration and risk of multiple sclerosis. *Neurology*, 28 (11), 1089–1089.
- Alter, M., Kahana, E., Zilber, N., and Miller, A., 2006. Multiple sclerosis frequency in Israel's diverse populations. *Neurology*, 66 (7), 1061–1066.
- Alter, M., Leibowitz, U., and Speer, J., 1966. Risk of multiple sclerosis related to age at immigration to Israel. *Archives of Neurology*, 15 (3), 234–237.
- Alter, M., Loewenson, R., and Harshe, M., 1973. The geographic distribution of multiple sclerosis: An examination of mathematical models. *Journal of Chronic Diseases*, 26 (12), 755–767.
- Antonovsky, A., Leibowitz, U., Smith, H. A., Medalie, J. M., Balogh, M., Kats, R., Halpern, L., and Alter, M., 1965. Epidemiologic study of multiple sclerosis in Israel: I. An overall review of methods and findings. *Archives of Neurology*, 13 (2), 183–193.
- Ares, B., Prieto, J. M., Lema, M., Dapena, D., Arias, M., and Noya, M., 2007. Prevalence of multiple sclerosis in Santiago de Compostela (Galicia, Spain). *Multiple Sclerosis*, 13 (2), 262–264.
- Argyriadou, S., Makridis, D., Lygidakis, H., Apazidis, G., and Gagalis, G., 2005. Knowledge and behaviour of tourists towards the sun, as studied in a region of northern Greece. *Rural and Remote Health*, 5 (4), 367.
- Arnason, B. G. W., 2011. MS forum/MS over the past 17 years. *International MS Journal*, 17 (3), 76–82.

- Arnold, B. C., and Beaver, R. J., 2000. Hidden truncation models. *Sankhyā: The Indian Journal of Statistics, Series A*, 62 (1), 23–35.
- Arnold, B. C., and Beaver, R. J., 2002. Skewed multivariate models related to hidden truncation and/or selective reporting. *Test*, 11 (1), 7–54.
- Arnold, B. C., Beaver, R. J., Groeneveld, R. A., and Meeker, W. Q., 1993. The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika*, 58 (3), 471–488.
- Artuković, M., Ikić, M., Kuštelega, J., Artuković, I. N., and Kaliterna, D. M., 2010. Influence of UV radiation on immunological system and occurrence of autoimmune diseases. *Collegium Antropologicum*, 34 (Suppl. 2), 175–178.
- Arunabh, S., Pollack, S., Yeh, J., and Aloia, J. F., 2003. Body fat content and 25-hydroxyvitamin D levels in healthy women. *Journal of Clinical Endocrinology and Metabolism*, 88 (1), 157–161.
- Ascherio, A., and Munger, K. L., 2007a. Environmental risk factors for multiple sclerosis. Part I: The role of infection. *Annals of Neurology*, 61 (4), 288–299.
- Ascherio, A., and Munger, K. L., 2007b. Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors. *Annals of Neurology*, 61 (6), 504–513.
- Ascherio, A., Munger, K. L., and Simon, K. C., 2010. Vitamin D and multiple sclerosis. *Lancet Neurology*, 9 (6), 599–612.
- AUAIHW. 2011. Melanoma of the skin [Microsoft Excel spreadsheet]: Australian Cancer Incidence and Mortality (ACIM) books. Available from: <http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=10737420938&libID=10737420937> [Accessed: 2012].
- Aubrecht, C., Steinnocher, K., Hollaus, M., and Wagner, W., 2009. Integrating earth observation and GIScience for high resolution spatial and functional modeling of urban land use. *Computers, Environment and Urban Systems*, 33 (1), 15–25.
- Auckland City Libraries. n.d. Auckland City Maps Online Available from: <http://www.aucklandcity.govt.nz/dbtw-wpd/maps/maps.html> [Accessed: 2012]
- AUDHA. 2010. National skin cancer awareness campaign Available from: <http://www.skincancer.gov.au/internet/skincancer/publishing.nsf/Content/home> [Accessed: 2012]

- Autier, P., Boniol, M., and Doré, J.-F., 2007. Sunscreen use and increased duration of intentional sun exposure: Still a burning issue. *International Journal of Cancer*, 121 (1), 1–5.
- Azzalini, A., 1985. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12 (2), 171–178.
- Azzalini, A., 2010. R package 'sn': The skew-normal and skew-t distributions (version 0.4-16). Università di Padova, Italia. Available from: <http://azzalini.stat.unipd.it/SN> [Accessed 2012].
- Azzalini, A., and Dalla Valle, A., 1996. The multivariate skew-normal distribution. *Biometrika*, 83 (4), 715–726.
- Bains, W., 2010. Exposure of the eyes to near-horizon sunshine may be a trigger for multiple sclerosis. *Medical Hypotheses*, 74 (3), 428–432.
- Bakhtiarova, K. Z., and Magzhanov, R. V., 2006. Рассеянный склероз в этнических группах республики башкортостан [Multiple sclerosis in ethnic groups of Bashkortostan Republic]. *Zhurnal Nevrologii i Psikhatrii imeni S.S. Korsakova*, Spec No 3, 17–21 (in Russian).
- Bălașa, R., Feier, C., Dan, M., Motataianu, A., Balaianu, M., Chebut, C., Constantin, V., Bajko, Z., and Pascu, I., 2007. The prevalence of multiple sclerosis in Mures county, central Romania. *Revista Română de Neurologie*, 6 (2), 80–84.
- Bar-Or, A., Oliveira, E. M. L., Anderson, D. E., and Hafler, D. A., 1999. Molecular pathogenesis of multiple sclerosis. *Journal of Neuroimmunology*, 100 (1–2), 252–259.
- Baranzini, S. E., 2011. Revealing the genetic basis of multiple sclerosis: Are we there yet? *Current Opinion in Genetics & Development*, 21 (3), 317–324.
- Barnett, M. H., Parratt, J. D. E., Pollard, J. D., and Prineas, J. W., 2009. MS: Is it one disease? *International MS Journal*, 16 (2), 57–65.
- Barnett, M. H., Williams, D. B., Day, S., Macaskill, P., and McLeod, J. G., 2003. Progressive increase in incidence and prevalence of multiple sclerosis in Newcastle, Australia: A 35-year study. *Journal of the Neurological Sciences*, 213 (1–2), 1–6.
- Baumhackl, U., Eibl, G., Ganzinger, U., Hartung, H. P., Mamoli, B., Pfeiffer, K. P., Fazekas, F., and Vass, K., 2002. Prevalence of multiple sclerosis in Austria: Results of a nationwide survey. *Neuroepidemiology*, 21 (5), 226–234.

- Bayes, H. K., Weir, C. J., and O'Leary, C., 2010. Timing of birth and risk of multiple sclerosis in the Scottish population. *European Neurology*, 63 (1), 36–40.
- Becker, R. A., and Chambers, J. M., 1984. *S. An interactive environment for data analysis and graphics*. Monterey: Wadsworth and Brooks/Cole.
- Becker, R. A., Chambers, J. M., and Wilks, A. R., 1988. *The new S language*. London: Chapman & Hall.
- Behan, P., and Chaudhuri, A., 2010. The sad plight of multiple sclerosis research (low on fact, high on fiction): Critical data to support it being a neurocristopathy. *Inflammopharmacology*, 18 (6), 265–290.
- Behan, P. O., 2010. Futility of the autoimmune orthodoxy in multiple sclerosis research. *Expert Review of Neurotherapeutics*, 10 (7), 1023–1025.
- Benamer, H. T. S., Ahmed, E. S. M., Al-Din, A. S., and Grosset, D. G., 2009. Frequency and clinical patterns of multiple sclerosis in Arab countries: A systematic review. *Journal of the Neurological Sciences*, 278 (1–2), 1–4.
- Bencsik, K., Rajda, C., Füvesi, J., Klivényi, P., Járdánházy, T., Török, M., and Vécsei, L., 2001. The prevalence of multiple sclerosis, distribution of clinical forms of the disease and functional status of patients in Csongrád county, Hungary. *European Neurology*, 46 (4), 206–209.
- Bentham, G., 1988. Migration and morbidity: Implications for geographical studies of disease. *Social Science & Medicine*, 26 (1), 49–54.
- Bentzen, J., Meulengracht Flachs, E., Stenager, E., Brønnum-Hansen, H., and Koch-Henriksen, N., 2010. Prevalence of multiple sclerosis in Denmark 1950–2005. *Multiple Sclerosis*, 16 (5), 520–525.
- Beretich, B. D., and Beretich, T. M., 2009. Explaining multiple sclerosis prevalence by ultraviolet exposure: A geospatial analysis. *Multiple Sclerosis*, 15 (8), 891–898.
- Bettini, M., and Vignali, D. A. A., 2009. Regulatory T cells and inhibitory cytokines in autoimmunity. *Current Opinion in Immunology*, 21 (6), 612–618.
- Bhigjee, A. I., Moodley, K., and Ramkissoon, K., 2007. Multiple sclerosis in KwaZulu Natal, South Africa: An epidemiological and clinical study. *Multiple Sclerosis*, 13 (9), 1095–1099.
- Binzer, S., Imrell, K., Binzer, M., Vang, S., Rogvi-Hansen, B., Hillert, J., and Stenager, E., 2010. Multiple sclerosis in a family on the Faroe Islands. *Acta Neurologica Scandinavica*, 121 (1), 16–19.

- Binzer, S., Imrell, K., Binzer, M., Vang, S., Rogvi-Hansen, B., Hillert, J., and Stenager, E., 2011. Multiple sclerosis in a family on the Faroe Islands: Reply. *Acta Neurologica Scandinavica*, 121 (6), 432.
- Björk, J., and Strömberg, U., 2002. Effects of systematic exposure assessment errors in partially ecologic case-control studies. *International Journal of Epidemiology*, 31 (1), 154–160.
- Björk, J., and Strömberg, U., 2005. Model specification and unmeasured confounders in partially ecologic analyses based on group proportions of exposed. *Scandinavian Journal of Work, Environment & Health*, 31 (3), 184-190.
- Bland, M., 2000. *An introduction to medical statistics*. Third ed.: Oxford University Press.
- Blum, M., Dolnikowski, G., Seyoum, E., Harris, S., Booth, S., Peterson, J., Saltzman, E., and Dawson-Hughes, B., 2008. Vitamin D₃ in fat tissue. *Endocrine*, 33 (1), 90–94.
- Bogh, M. K. B., Schmedes, A. V., Philipsen, P. A., Thieden, E., and Wulf, H. C., 2011. Interdependence between body surface area and ultraviolet B dose in vitamin D production: A randomized controlled trial. *British Journal of Dermatology*, 164 (1), 163–169.
- Bolande, R. P., 1974. The neurocristopathies: A unifying concept of disease arising in neural crest maldevelopment. *Human Pathology*, 5 (4), 409–429.
- Bolland, M. J., Grey, A., Cundy, T., and Reid, I. R., 2007a. Defining vitamin D deficiency. *New Zealand Medical Journal*, 120 (1263), <http://www.nzma.org.nz/journal/120-1263/2760/>.
- Bolland, M. J., Grey, A. B., Ames, R. W., Mason, B. H., Horne, A. M., Gamble, G. D., and Reid, I. R., 2007b. The effects of seasonal variation of 25-hydroxyvitamin D and fat mass on a diagnosis of vitamin D sufficiency. *American Journal of Clinical Nutrition*, 86 (4), 959–964.
- Boppana, S., Huang, H., Ito, K., and Dhib-Jalbut, S., 2011. Immunologic aspects of multiple sclerosis. *Mount Sinai Journal of Medicine*, 78 (2), 207–220.
- Börü, Ü. T., Alp, R., Sur, H., and Gül, L., 2006. Prevalence of multiple sclerosis door-to-door survey in Maltepe, Istanbul, Turkey. *Neuroepidemiology*, 27 (1), 17–21.

- Boström, I., Callander, M., Kurtzke, J. F., and Landtblom, A. M., 2009. High prevalence of multiple sclerosis in the Swedish county of Värmland. *Multiple Sclerosis*, 15 (11), 1253–1262.
- Brennan, R. M., Burrows, J. M., Bell, M. J., Bromham, L., Csurhes, P. A., Lenarczyk, A., Sverndal, J., Klintenstedt, J., Pender, M. P., and Burrows, S. R., 2010. Strains of Epstein–Barr virus infecting multiple sclerosis patients. *Multiple Sclerosis*, 16 (6), 643–651.
- Brindley, P., Wise, S. M., Maheswaran, R., and Haining, R. P., 2005. The effect of alternative representations of population location on the areal interpolation of air pollution exposure. *Computers, Environment and Urban Systems*, 29 (4), 455–469.
- Brookes, A. J., 1999. The essence of SNPs. *Gene*, 234 (2), 177–186.
- Brox, J., Bjørnstad, E., and Olausson, K., 2003. Hemoglobin, iron, nutrition and life-style among adolescents in a coastal and an inland community in northern Norway. *International Journal of Circumpolar Health*, 62 (2), 130–141.
- Brustad, M., Alsaker, E., Engelsen, O., Aksnes, L., and Lund, E., 2004a. Vitamin D status of middle-aged women at 65–71 °N in relation to dietary intake and exposure to ultraviolet radiation. *Public Health Nutrition*, 7 (2), 327–35.
- Brustad, M., Sandanger, T., Aksnes, L., and Lund, E., 2004b. Vitamin D status in a rural population of northern Norway with high fish liver consumption. *Public Health Nutrition*, 7 (6), 783–789.
- Brustad, M., Sandanger, T., Wilsgaard, T., Aksnes, L., and Lund, E., 2003. Change in plasma levels of vitamin D after consumption of cod-liver and fresh cod-liver oil as part of the traditional north norwegian fish dish "Mølje". *International Journal of Circumpolar Health*, 62 (1), 40–53.
- Burgaz, A., Åkesson, A., Öster, A., Michaëlsson, K., and Wolk, A., 2007. Associations of diet, supplement use, and ultraviolet B radiation exposure with vitamin D status in Swedish women during winter. *American Journal of Clinical Nutrition*, 86 (5), 1399–1404.
- Cabre, P., 2007. Migration and multiple sclerosis: The French West Indies experience. *Journal of the Neurological Sciences*, 262 (1–2), 117–121.

- Cabre, P., Signate, A., Olindo, S., Merle, H., Caparros-Lefebvre, D., Béra, O., and Smadja, D., 2005. Role of return migration in the emergence of multiple sclerosis in the French West Indies. *Brain*, 128 (12), 2899–2910.
- Callegaro, D., Goldbaum, M., Morais, L., Tilbery, C. P., Moreira, M. A., Gabbai, A. A., and Scaff, M., 2001. The prevalence of multiple sclerosis in the city of São Paulo, Brazil, 1997. *Acta Neurologica Scandinavica*, 104 (4), 208–213.
- Calvo, M. S., Whiting, S. J., and Barton, C. N., 2005. Vitamin D intake: A global perspective of current status. *Journal of Nutrition*, 135 (2), 310–316.
- Carolei, A., and Sacco, S., 2011. Multiple sclerosis in the Faroe Islands: A core context for neuroanthropology. *Acta Neurologica Scandinavica*, 124 (1), 64–65.
- Casquero, P., Villoslada, P., Montalbán, X., and Torrent, M., 2001. Frequency of multiple sclerosis in Menorca, Balearic Islands, Spain. *Neuroepidemiology*, 20 (2), 129–133.
- CCA. 2012. SunSmart Available from:
<http://www.cancer.org.au/cancersmartlifestyle/SunSmart.htm> [Accessed: 2012]
- Chambers, J. M., 1998. *Programming with data*. New York: Springer.
- Chambers, J. M., and Hastie, T. J., 1992. *Statistical models in S*. London: Chapman & Hall.
- Chastain, E. M. L., Duncan, D. A. S., Rodgers, J. M., and Miller, S. D., 2011. The role of antigen presenting cells in multiple sclerosis. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1812 (2), 265–274.
- Chen, T. C., Chimeh, F., Lu, Z., Mathieu, J., Person, K. S., Zhang, A., Kohn, N., Martinello, S., Berkowitz, R., and Holick, M. F., 2007. Factors that influence the cutaneous synthesis and dietary sources of vitamin D. *Archives of Biochemistry and Biophysics*, 460 (2), 213–217.
- Chen, T. C., Lu, Z., and Holick, M. F., 2010. Photobiology of vitamin D. *In*: Holick, M. F. ed. *Vitamin D*: Springer Science and Business Media, 35–60.
- Christchurch City Libraries. n.d. Heritage Maps Available from:
<http://christchurchcitylibraries.com/Heritage/Maps> [Accessed: 2012]
- Christensen, T., 2006. The role of EBV in MS pathogenesis. *International MS Journal*, 13, 53–57.

- CIE. 2006. *Action spectrum for the production of previtamin D₃ in human skin*. (Technical Report CIE 174:2006). Vienna, Austria: Commission Internationale de l'Eclairage (CIE) Central Bureau.
- Clemens, T. L., Henderson, S. L., Adams, J. S., and Holick, M. F., 1982. Increased skin pigment reduces the capacity of skin to synthesise vitamin D₃. *Lancet*, 319 (8263), 74–76.
- Coggan, J. S., Prescott, S. A., Bartol, T. M., and Sejnowski, T. J., 2010. Imbalance of ionic conductances contributes to diverse symptoms of demyelination. *Proceedings of the National Academy of Sciences*, 107 (48), 20602–20609.
- Cohen, J. E., Roig, M., Reuman, D. C., and Gogwilt, C., 2008. International migration beyond gravity: A statistical model for use in population projections. *Proceedings of the National Academy of Sciences*, 105 (40), 15269–15274.
- Compston, A., and Coles, A., 2008. Multiple sclerosis. *Lancet*, 372 (9648), 1502–1517.
- Corona, T., and Román, G. C., 2006. Multiple sclerosis in Latin America. *Neuroepidemiology*, 26, 1–3.
- Correale, J., Ysraelit, M. C., and Gaitán, M. I., 2009. Immunomodulatory effects of vitamin D in multiple sclerosis. *Brain*, 132 (5), 1146–1160.
- Cristiano, E., Patrucco, L., and Rojas, J. I., 2008. A systematic review of the epidemiology of multiple sclerosis in South America. *European Journal of Neurology*, 15 (12), 1273–1278.
- Cristiano, E., Patrucco, L., Rojas, J. I., Cáceres, F., Carrá, A., Correale, J., Garcea, O., Gold, L., Tessler, J., and Kremenchutzky, M., 2009. Prevalence of multiple sclerosis in Buenos Aires, Argentina using the capture-recapture method. *European Journal of Neurology*, 16 (2), 183–187.
- Cromley, R. G., Ebenstein, A. Y., and Hanink, D. M., 2009. Estimating components of population change from census data for incongruent spatial/temporal units and attributes. *Journal of Spatial Science*, 54, 89–99.
- Cross, A. H., and Parks, B. J., 2010. The four seasons of multiple sclerosis. *Neurology*, 75 (9), 762–763.
- CRUK. 2007. Sunburnt holiday-makers boost their chances of fatal skin cancer *News & Resources* Available from: <http://info.cancerresearchuk.org/news/archive/pressrelease/2007-04-03->

- sunburnt-holidaymakers-boost-their-chances-of-fatal-skin-cancer [Accessed: 2012]
- CRUK. 2008. Holiday 'binge tanning' increasing skin cancer risk for young brits *News & Resources* Available from:
<http://info.cancerresearchuk.org/news/archive/pressrelease/2008-05-05-holiday-binge-tanning-increasing-skin-cancer-risk-for-young-brits> [Accessed: 2012]
- CRUK. 2009. Skin cancer prevention Available from:
<http://cancerhelp.cancerresearchuk.org/type/skin-cancer/about/preventing-skin-cancer> [Accessed: 2012]
- CRUK. 2012. SunSmart Available from: <http://www.sunsmart.org.uk/> [Accessed: 2012]
- Czarnecki, D., Meehan, C. J., and Bruce, F., 2009. The vitamin D status of Australian dermatologists. *Clinical and Experimental Dermatology*, 34 (5), 624–625.
- Dahl, O. P., Aarseth, J. H., Myhr, K. M., Nyland, H., and Midgard, R., 2004. Multiple sclerosis in Nord-Trøndelag County, Norway: A prevalence and incidence study. *Acta Neurologica Scandinavica*, 109 (6), 378–384.
- de Lourdes Samaniego-Vaesken, M., Alonso-Apperte, E., and Varela-Moreiras, G., 2012. Vitamin food fortification today. *Food & Nutrition Research*, 56, 5459–5467.
- De Sá, J., Paulos, A., Mendes, H., Becho, J., Marques, J., and Roxo, J., 2006. The prevalence of multiple sclerosis in the District of Santarém, Portugal. *Journal of Neurology*, 253 (7), 914–918.
- Dean, G., Elian, M., de Bono, A. G., Asciak, R. P., Vella, N., Mifsud, V., and Aquilina, J., 2002. Multiple sclerosis in Malta in 1999: An update. *Journal of Neurology, Neurosurgery and Psychiatry*, 73 (3), 256–260.
- Deutch, B., Dyerberg, J., Pedersen, H. S., Aschlund, E., and Hansen, J. C., 2007. Traditional and modern Greenlandic food — dietary composition, nutrients and contaminants. *Science of The Total Environment*, 384 (1–3), 106–119.
- Díaz, V., Barahona, J., Antinao, J., Quezada, R., Delgado, I., Silva, C., and Guiloff, R. J., 2011. Incidence of multiple sclerosis in Chile. A hospital registry study. *Acta Neurologica Scandinavica*, DOI: 10.1111/j.1600-0404.2011.01571.x.
- Diffey, B., 1998. Sun protection: Have we gone too far? *British Journal of Dermatology*, 138 (3), 562–563.

- Diffey, B., 2006. Do we need a revised public health policy on sun exposure? *British Journal of Dermatology*, 154 (6), 1046–1051.
- Disanto, G., Meier, U., Giovannoni, G., and Ramagopalan, S. V., 2011. Vitamin D: A link between Epstein–Barr virus and multiple sclerosis development? *Expert Review of Neurotherapeutics*, 11 (9), 1221–1224.
- Diveu, C., McGeachy, M. J., and Cua, D. J., 2008. Cytokines that regulate autoimmunity. *Current Opinion in Immunology*, 20 (6), 663–668.
- DKMFLF. 2011. Ansøgninger om berigelse af produkter [Applications for the enrichment of products]. Available from: <http://www.foedevarestyrelsen.dk/Foedevarer/Ernaering/Berigelse/Sider/Berigningsansoegninger.aspx> [Accessed: 2012] (in Danish).
- DKMFLF. 2012. Generelle tilladelser til tilsætning af vitaminer og mineraler [General permits for the addition of vitamins and minerals]. Available from: http://www.foedevarestyrelsen.dk/Foedevarer/Ernaering/Berigelse/Sider/Generelle_tilladelser_til_berigelse.aspx [Accessed: 2012] (in Danish).
- DKMFLF. 2012. Liste over ansøgninger om berigelse af fødevarer [List of applications for food fortification]. Available from: http://www.foedevarestyrelsen.dk/SiteCollectionDocuments/25_PDF_word_fil%20til%20download/07kontor/Maerkning/Liste%20over%20berigelsessager%20til%20offentlig%20B%20relse%20p%20A%20hjemmesiden.doc [Accessed: 2012] (in Danish).
- Dobbinson, S. J., Wakefield, M. A., Jansen, K. M., Herd, N. L., Spittal, M. J., Lipscomb, J. E., and Hill, D. J., 2008. Weekend sun protection and sunburn in Australia: Trends (1987–2002) and association with SunSmart television advertising. *American Journal of Preventive Medicine*, 34 (2), 94–101.
- Dobec-Meić, B., Puljić, I., and Kalousek, M., 2008. Prevalence of multiple sclerosis in Varaždin region. *Neurologia Croatica*, 57 (1–2), 3–9.
- Dobnig, H., 2011. A review of the health consequences of the vitamin D deficiency pandemic. *Journal of the Neurological Sciences*, 311 (1–2), 15–18.
- Dokuchaeva, N. N., and Boïko, A. N., 2006. Клинико-эпидемиологическое исследование рассеянного склероза в волгограде [Clinical and epidemiological study of multiple sclerosis in Volgograd city]. *Zhurnal Nevrologii i Psikiatrii imeni S.S. Korsakova, Spec No 3*, 4–10 (in Russian).

- Doll, R., and Cook, P., 1967. Summarizing indices for comparison of cancer incidence data. *International Journal of Cancer*, 2, 269–279.
- Doll, R., Payne, P. M., and Waterhouse, J. A. H., 1966. *Cancer incidence in five countries*. International Union Against Cancer, Springer-Verlag, Berlin.
- Drulović, J., Mostarica-Stojković, M., Lević, Z., Stojsavljević, N., Pravica, V., and Mesaroš, Š., 1997. Interleukin-12 and tumor necrosis factor- α levels in cerebrospinal fluid of multiple sclerosis patients. *Journal of the Neurological Sciences*, 147 (2), 145–150.
- Dutta, R., and Trapp, B. D., 2011. Mechanisms of neuronal dysfunction and degeneration in multiple sclerosis. *Progress in Neurobiology*, 93 (1), 1–12.
- Dyment, D. A., Ebers, G. C., and Sadovnick, A. D., 2004. Genetics of multiple sclerosis. *Lancet Neurology*, 3 (2), 104–110.
- Ebers, G., and Goodin, D., 2007. Multiple sclerosis: Epidemiology, genetics and environmental factors. *Proceedings of the MS Forum Modern Management Workshop, Wiesbaden, February 2007*.
- Ebers, G. C., 2008. Environmental factors and multiple sclerosis. *Lancet Neurology*, 7 (3), 268–277.
- Ebers, G. C., 2009. The emperor's new clothiers. *International MS Journal*, 16 (1), 3–4.
- Ebers, G. C., and Sadovnick, A. D., 1993. The geographic distribution of multiple sclerosis: A review. *Neuroepidemiology*, 12 (1), 1–5.
- Edvardsen, K., Brustad, M., Engelsen, O., and Aksnes, L., 2007. The solar UV radiation level needed for cutaneous production of vitamin D₃ in the face. A study conducted among subjects living at a high latitude (68° N). *Photochemical & Photobiological Sciences*, 6 (1), 57–62.
- Eikelenboom, M. J., Killestein, J., Kragt, J. J., Uitdehaag, B. M. J., and Polman, C. H., 2009. Gender differences in multiple sclerosis: Cytokines and vitamin D. *Journal of the Neurological Sciences*, 286 (1–2), 40–42.
- El-Etr, M., Ghomari, A., Sitruk-Ware, R., and Schumacher, M., 2011. Hormonal influences in multiple sclerosis: New therapeutic benefits for steroids. *Maturitas*, 68 (1), 47–51.
- Elhami, S. R., Mohammad, K., Sahraian, M. A., and Eftekhari, H., 2011. A 20-year incidence trend (1989–2008) and point prevalence (March 20, 2009) of multiple

- sclerosis in Tehran, Iran: A population-based study. *Neuroepidemiology*, 36 (3), 141–147.
- Elian, M., Nightingale, S., and Dean, G., 1990. Multiple sclerosis among United Kingdom-born children of immigrants from the Indian subcontinent, Africa and the West Indies. *Journal of Neurology, Neurosurgery and Psychiatry*, 53 (10), 906–911.
- Engelsen, O., 2010. FastRT Available from: <http://nadir.nilu.no/~olaeng/fastrt/fastrt.html> [Accessed: 2012]
- Engelsen, O., 2010b. The relationship between ultraviolet radiation exposure and vitamin D status. *Nutrients*, 2 (5), 482–495.
- Engelsen, O., Brustad, M., Aksnes, L., and Lund, E., 2005. Daily duration of vitamin D synthesis in human skin with relation to latitude, total ozone, altitude, ground cover, aerosols and cloud thickness. *Photochemistry and Photobiology*, 81, 1287–1290.
- Engelsen, O., and Kylling, A., 2005. Fast simulation tool for ultraviolet radiation at the earth's surface. *Optical Engineering*, 44 (4), 041042.
- ESRI. 2009. ArcGIS (Version 9.3.1): ESRI. Available from: <http://www.esri.com>.
- Etemadifar, M., Janghorbani, M., Shaygannejad, V., and Ashtari, F., 2006. Prevalence of multiple sclerosis in Isfahan, Iran. *Neuroepidemiology*, 27 (1), 39–44.
- Ezzedine, K., Guinot, C., Mauger, E., Pistone, T., Receveur, M.-C., Galan, P., Hercberg, S., and Malvy, D., 2007. Travellers to high UV-index countries: Sun-exposure behaviour in 7822 French adults. *Travel Medicine and Infectious Disease*, 5 (3), 176–182.
- Fatima, N., Toscano, M. P., Hunter, S. B., and Cohen, C., 2011. Controversial role of Epstein–Barr virus in multiple sclerosis. *Applied Immunohistochemistry & Molecular Morphology*, 19 (3), 246–252 10.1097/PAI.0b013e3181f3b4.
- Fernández, O., Fernández, V., Martínez-Cabrera, V., Mayorga, C., Alonso, A., León, A., Arnal, C., Hens, M., Luque, G., De Ramón, E., Caballero, A., and Leyva, L., 2008. Multiple sclerosis in Gypsies from southern Spain: Prevalence, mitochondrial DNA haplogroups and HLA class II association. *Tissue Antigens*, 71 (5), 426–433.
- Firth, D., 1941. The case of Augustus d'Esté (1794–1848): The first account of disseminated sclerosis. *Proceedings of the Royal Society of Medicine*, 34, 381–384.

- Fletcher, R. J., Bell, I. P., and Lambert, J. P., 2004. Public health aspects of food fortification: A question of balance. *Proceedings of the Nutrition Society*, 63 (4), 605–614.
- Flowerdew, R., and Green, M., 1992. Developments in areal interpolation methods and GIS. *Annals of Regional Science*, 26 (1), 67–78.
- Flowerdew, R., Green, M., and Kehris, E., 1991. Using areal interpolation methods in geographic information systems. *Papers in Regional Science*, 70 (3), 303–315.
- Flowerdew, R., Manley, D. J., and Sabel, C. E., 2008. Neighbourhood effects on health: Does it matter where you draw the boundaries? *Social Science & Medicine*, 66 (6), 1241–1255.
- Foss, Y. J., 2009. Vitamin D deficiency is the cause of common obesity. *Medical Hypotheses*, 72 (3), 314–321.
- Fox, C. M., Bensa, S., Bray, I., and Zajicek, J. P., 2004. The epidemiology of multiple sclerosis in Devon: A comparison of the new and old classification criteria. *Journal of Neurology, Neurosurgery and Psychiatry*, 75 (1), 56–60.
- Fragoso, Y. D., and Peres, M., 2007. Prevalence of multiple sclerosis in the city of Santos, SP. *Revista Brasileira de Epidemiologia*, 10 (4), 479–482.
- Fromont, A., Binquet, C., Sauleau, E. A., Fournel, I., Bellisario, A., Adnet, J., Weill, A., Vukusic, S., Confavreux, C., Debouverie, M., Clerc, L., Bonithon-Kopp, C., and Moreau, T., 2010. Geographic variations of multiple sclerosis in France. *Brain*, 133 (7), 1889–1899.
- Gandhi, R., Laroni, A., and Weiner, H. L., 2010. Role of the innate immune system in the pathogenesis of multiple sclerosis. *Journal of Neuroimmunology*, 221 (1–2), 7–14.
- Gautier, L., 2009. rpy2: Python interface to the R language (Version 2.0.6). Available from: <http://rpy.sourceforge.net/rpy2.html> [Accessed 2012].
- Ghandehari, K., Riasi, H. R., Nourian, A., and Boroumand, A. R., 2010. Prevalence of multiple sclerosis in north east of Iran. *Multiple Sclerosis*, 16 (12), 1525–1526.
- Gibbs, N. K., and Norval, M., 2011. Urocanic acid in the skin: A mixed blessing? *Journal of Investigative Dermatology*, 131 (1), 14–17.
- Gilaberte, Y., Aguilera, J., Carrascosa, J. M., Figueroa, F. L., Romaní De Gabriel, J., and Nagore, E., 2011. Vitamin D: Evidence and controversies. *Actas Dermo-Sifiliográficas*, 102 (8), 572–588.

- Gillie, O., 2006. A new government policy is needed for sunlight and vitamin D. *British Journal of Dermatology*, 154 (6), 1052–1061.
- Gillie, O., 2010. Sunlight robbery: A critique of public health policy on vitamin D in the UK. *Molecular Nutrition & Food Research*, 54 (8), 1148–1163.
- Giovannoni, G., 2011. Epstein–Barr virus and MS. *International MS Journal*, 17 (2), 44–49.
- Glass, D., Lens, M., Swaminathan, R., Spector, T. D., and Bataille, V., 2009. Pigmentation and vitamin D metabolism in caucasians: Low vitamin D serum levels in fair skin types in the UK. *PLoS ONE*, 4 (8), e6477.
- Glerup, H., Mikkelsen, K., Poulsen, L., Hass, E., Overbeck, S., Thomsen, J., Charles, P., and Eriksen, E. F., 2000. Commonly recommended daily intake of vitamin D is not sufficient if sunlight exposure is limited. *Journal of Internal Medicine*, 247, 260–268.
- Gong, X., Xie, Z., and Zuo, H., 2008. A new track for understanding the pathogenesis of multiple sclerosis: From the perspective of early developmental deficit caused by the potential 5-HT deficiency in individuals in high-latitude areas. *Medical Hypotheses*, 71 (4), 580–583.
- Goodchild, M. F., Anselin, L., and Deichmann, U., 1993. A framework for the areal interpolation of socioeconomic data. *Environment and Planning A*, 25, 383–397.
- Goodchild, M. F., and Lam, N. S.-N., 1980. Areal interpolation: A variant of the traditional spatial problem. *Geo-Processing*, 1, 297–312.
- Goodin, D., 2010. The genetic basis of multiple sclerosis: A model for MS susceptibility. *BMC Neurology*, 10, 101.
- Goodin, D. S., 2009. The causal cascade to multiple sclerosis: A model for MS pathogenesis. *PLoS ONE*, 4 (2), e4565.
- Gotway, C. A., and Young, L. J., 2008. A geostatistical approach to linking geographically aggregated data from different sources. *Journal of Computational and Graphical Statistics*, 16 (1), 115–135.
- Gourraud, P.-A., Harbo, H. F., Hauser, S. L., and Baranzini, S. E., 2012. The genetics of multiple sclerosis: An up-to-date review. *Immunological Reviews*, 248 (1), 87–103.

- Gozdzik, A., Barta, J. L., Wu, H., Wagner, D., Cole, D. E., Vieth, R., Whiting, S., and Parra, E. J., 2008. Low wintertime vitamin D levels in a sample of healthy young adults of diverse ancestry living in the Toronto area: Associations with vitamin D intake and skin pigmentation. *BMC Public Health*, 8, 336.
- Gracia, F., Castillo, L. C., Benzaón, A., Larreategui, M., Villareal, F., Triana, E., Arango, A. C., Lee, D., Pascale, J. M., Gomez, E., and Armien, B., 2009. Prevalence and incidence of multiple sclerosis in Panama (2000–2005). *Neuroepidemiology*, 32 (4), 287–293.
- Granieri, E., Economou, N. T., De Gennaro, R., Tola, M. R., Caniatti, L., Govoni, V., Fainardi, E., and Casetta, I., 2007. Multiple sclerosis in the province of Ferrara: Evidence for an increasing trend. *Journal of Neurology*, 254 (12), 1642–1648.
- Grant, W. B., and Holick, M. F., 2005. Benefits and requirements of vitamin D for optimal health: A review. *Alternative Medicine Review*, 10 (2), 94–111.
- Gray, O. M., McDonnell, G. V., and Hawkins, S. A., 2008. Factors in the rising prevalence of multiple sclerosis in the north-east of Ireland. *Multiple Sclerosis*, 14 (7), 880–886.
- Green, J., 1951. Studies on the irradiation of ergosterol and 7-dehydrocholesterol and the analysis of the products for calciferol, vitamin D₃, and component sterols. *Biochemical Journal*, 49 (2), 232–243.
- Greenland, S., 2001. Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. *International Journal of Epidemiology*, 30 (6), 1343–1350.
- Greer, J. M., and McCombe, P. A., 2011. Role of gender in multiple sclerosis: Clinical effects and potential molecular mechanisms. *Journal of Neuroimmunology*, 234 (1–2), 7–18.
- Gregory, I. N., 2002. The accuracy of areal interpolation techniques: Standardising 19th and 20th century census data to allow long-term comparisons. *Computers, Environment and Urban Systems*, 26 (4), 293–314.
- Gregory, I. N., and Ell, P. S., 2005. Breaking the boundaries: Geographical approaches to integrating 200 years of the census. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 168 (2), 419–437.

- Gregory, I. N., and Ell, P. S., 2006. Error-sensitive historical GIS: Identifying areal interpolation errors in time-series data. *International Journal of Geographical Information Science*, 20 (2), 135–152.
- Grimaldi, L. M. E., Palmeri, B., Salemi, G., Giglia, G., D'Amelio, M., Grimaldi, R., Vitello, G., Ragonese, P., and Savettieri, G., 2007. High prevalence and fast rising incidence of multiple sclerosis in Caltanissetta, Sicily, southern Italy. *Neuroepidemiology*, 28 (1), 28–32.
- Grønlie, S. A., Myrvoll, E., Hansen, G., Grønning, M., and Mellgren, S. I., 2000. Multiple sclerosis in North Norway, and first appearance in an indigenous population. *Journal of Neurology*, 247 (2), 129–133.
- Hader, W. J., and Yee, I. M. L., 2007. Incidence and prevalence of multiple sclerosis in Saskatoon, Saskatchewan. *Neurology*, 69 (12), 1224–1229.
- Halliday, G. M., 2010. Common links among the pathways leading to UV-induced immunosuppression. *Journal of Investigative Dermatology*, 130 (5), 1209–1212.
- Handel, A. E., Williamson, A. J., Disanto, G., Dobson, R., Giovannoni, G., and Ramagopalan, S. V., 2011. Smoking and multiple sclerosis: An updated meta-analysis. *PLoS ONE*, 6 (1), e16149.
- Handel, A. E., Williamson, A. J., Disanto, G., Handunnetthi, L., Giovannoni, G., and Ramagopalan, S. V., 2010. An updated meta-analysis of risk of multiple sclerosis following infectious mononucleosis. *PLoS ONE*, 5 (9), e12496.
- Handunnetthi, L., Ramagopalan, S. V., and Ebers, G. C., 2010. Multiple sclerosis, vitamin D, and HLA-DRB1*15. *Neurology*, 74 (23), 1905–1910.
- Hanwell, H. E. C., and Banwell, B., 2011. Assessment of evidence for a protective role of vitamin D in multiple sclerosis. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1812 (2), 202–212.
- Harbo, H. F., Utsi, E., Lorentzen, Å. R., Kampman, M. T., Celius, E. G., Myhr, K. M., Lie, B. A., Mellgren, S. I., and Thorsby, E., 2007. Low frequency of the disease-associated DRB1*15-DQB1*06 haplotype may contribute to the low prevalence of multiple sclerosis in Sami. *Tissue Antigens*, 69 (4), 299–304.
- Harris, R., and Chen, Z., 2005. Giving dimension to point locations: Urban density profiling using population surface models. *Computers, Environment and Urban Systems*, 29 (2), 115–132.

- Hashem, S., El-Tamawy, M., Hamdy, S., and Elmasry, T., 2010. Epidemiology of multiple sclerosis in Egypt. *Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, 47 (4), 625–632.
- Hassan, A. A., Sandanger, T. M., and Brustad, M., 2012. Level of selected nutrients in meat, liver, tallow and bone marrow from semi-domesticated reindeer (*Rangifer t. tarandus* L.). *International Journal of Circumpolar Health*, 71 (1), 17997 - DOI: 10.3402/ijch.v71i0.17997.
- Hauser, S. L., and Oksenberg, J. R., 2006. The neurobiology of multiple sclerosis: Genes, inflammation, and neurodegeneration. *Neuron*, 52 (1), 61–76.
- Hawkes, C. H., 2007. Smoking is a risk factor for multiple sclerosis: A metaanalysis. *Multiple Sclerosis*, 13 (5), 610–615.
- Heaney, R. P., Horst, R. L., Cullen, D. M., and Armas, L. A. G., 2009. Vitamin D₃ distribution and status in the body. *Journal of the American College of Nutrition*, 28 (3), 252–256.
- Heaney, R. P., Recker, R. R., Grote, J., Horst, R. L., and Armas, L. A. G., 2011. Vitamin D₃ is more potent than vitamin D₂ in humans. *Journal of Clinical Endocrinology and Metabolism*, 96 (3), E447–E452.
- Hedström, A. K., Bäärnhielm, M., Olsson, T., and Alfredsson, L., 2009. Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis. *Neurology*, 73 (9), 696–701.
- Hedström, A. K., Bäärnhielm, M., Olsson, T., and Alfredsson, L., 2011a. Exposure to environmental tobacco smoke is associated with increased risk for multiple sclerosis. *Multiple Sclerosis Journal*, 17 (7), 788–793.
- Hedström, A. K., Sundqvist, E., Bäärnhielm, M., Nordin, N., Hillert, J., Kockum, I., Olsson, T., and Alfredsson, L., 2011b. Smoking and two human leukocyte antigen genes interact to increase the risk for multiple sclerosis. *Brain*, 134 (3), 653–664.
- Hemmer, B., Nessler, S., Zhou, D., Kieseier, B., and Hartung, H.-P., 2006. Immunopathogenesis and immunotherapy of multiple sclerosis. *Nature Clinical Practice Neurology*, 2 (4), 201–211.
- Hepper, P., 2006. Prenatal development. *In: Slater, A., and Lewis, M. eds. Introduction to infant development: Oxford University Press*, 41–62.

- Hernán, M. A., Jick, S. S., Logroscino, G., Olek, M. J., Ascherio, A., and Jick, H., 2005. Cigarette smoking and the progression of multiple sclerosis. *Brain*, 128 (6), 1461–1465.
- Hernán, M. A., Oleky, M. J., and Ascherio, A., 2001. Cigarette smoking and incidence of multiple sclerosis. *American Journal of Epidemiology*, 154 (1), 69–74.
- Hiom, S., 2006. Public awareness regarding UV risks and vitamin D—the challenges for UK skin cancer prevention campaigns. *Progress in Biophysics and Molecular Biology*, 92 (1), 161–166.
- Hirst, C., Ingram, G., Pickersgill, T., Swingler, R., Compston, D. A. S., and Robertson, N. P., 2009. Increasing prevalence and incidence of multiple sclerosis in south east Wales. *Journal of Neurology, Neurosurgery and Psychiatry*, 80 (4), 386–391.
- Hoffjan, S., and Akkad, D. A., 2010. The genetics of multiple sclerosis: An update 2010. *Molecular and Cellular Probes*, 24 (5), 237–243.
- Holick, M. F., 1994. Vitamin D — new horizons for the 21st century. *American Journal of Clinical Nutrition*, 60, 619–630.
- Holick, M. F., 2004a. Sunlight and vitamin D for bone health and prevention of autoimmune diseases, cancers, and cardiovascular disease. *American Journal of Clinical Nutrition*, 80 (6), 1678S–1688S.
- Holick, M. F., 2004b. Vitamin D: Importance in the prevention of cancers, type 1 diabetes, heart disease, and osteoporosis. *American Journal of Clinical Nutrition*, 79 (3), 362–371.
- Holick, M. F., 2006. Resurrection of vitamin D deficiency and rickets. *Journal of Clinical Investigation*, 116 (8), 2062–2072.
- Holick, M. F., 2009. Vitamin D and health: Evolution, biologic functions, and recommended dietary intakes for vitamin D. *Clinical Reviews in Bone and Mineral Metabolism*, 7 (1), 2–19.
- Holick, M. F., 2010. The vitamin D deficiency pandemic: A forgotten hormone important for health. *Public Health Reviews*, 32 (1), 267–283.
- Holick, M. F., and Chen, T. C., 2008. Vitamin D deficiency: A worldwide problem with health consequences. *American Journal of Clinical Nutrition*, 87 (4), 1080S–1086S.
- Holick, M. F., Chen, T. C., Lu, Z., and Sauter, E., 2007. Vitamin D and skin physiology: A D-lightful story. *Journal of Bone and Mineral Research*, 22, V28–V33.

- Houzen, H., Niino, M., Hata, D., Nakano, F., Kikuchi, S., Fukazawa, T., and Sasaki, H., 2008. Increasing prevalence and incidence of multiple sclerosis in northern Japan. *Multiple Sclerosis*, 14 (7), 887–892.
- Houzen, H., Niino, M., Kikuchi, S., Fukazawa, T., Nogoshi, S., Matsumoto, H., and Tashiro, K., 2003. The prevalence and clinical characteristics of MS in northern Japan. *Journal of the Neurological Sciences*, 211 (1–2), 49–53.
- Hug, D. H., Hunter, J. K., and Dunkerson, D. D., 1998. The potential role for urocanic acid and sunlight in the immune suppression associated with protein malnutrition. *Journal of Photochemistry and Photobiology B: Biology*, 44, 117–123.
- Huotari, A., and Herzig, K. H., 2008. Vitamin D and living in northern latitudes — an endemic risk area for vitamin D deficiency. *International Journal of Circumpolar Health*, 67 (2–3), 164–178.
- Idris, M.-N. A., Sokrab, T. E. O., Ibrahim, E. A., Ali, H.-E.-R. S., Elzibair, M. A., Abadalatif, M., and Osman, R. R., 2009. Multiple sclerosis in Sudan: A prospective study of clinical presentation and outcome. *Multiple Sclerosis*, 15 (12), 1537–1538.
- Imitola, J., Chitnis, T., and Khoury, S. J., 2005. Cytokines in multiple sclerosis: From bench to bedside. *Pharmacology & Therapeutics*, 106 (2), 163–177.
- Imitola, J., Snyder, E. Y., and Khoury, S. J., 2003. Genetic programs and responses of neural stem/progenitor cells during demyelination: Potential insights into repair mechanisms in multiple sclerosis. *Physiological Genomics*, 14 (3), 171–197.
- IMSGC, and WTCCC. 2011. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*, 476 (7359), 214–219.
- Inshasi, J., and Thakre, M., 2011. Prevalence of multiple sclerosis in Dubai, United Arab Emirates. *International Journal of Neuroscience*, 121 (7), 393–398.
- Ishizu, T., Kira, J.-I., Osoegawa, M., Fukazawa, T., Kikuchi, S., Fujihara, K., Matsui, M., Kohriyama, T., Sobue, G., Yamamura, T., Itoyama, Y., Saida, T., and Sakata, K., 2009. Heterogeneity and continuum of multiple sclerosis phenotypes in Japanese according to the results of the fourth nationwide survey. *Journal of the Neurological Sciences*, 280 (1–2), 22–28.

- Islam, T., Gauderman, W. J., Cozen, W., Hamilton, A. S., Burnett, M. E., and Mack, T. M., 2006. Differential twin concordance for multiple sclerosis by latitude of birthplace. *Annals of Neurology*, 60 (1), 56–64.
- Itoh, T., Aizawa, H., Hashimoto, K., Kimura, T., Katayama, T., Koyama, S., Yahara, O., and Kikuchi, K., 2003. Prevalence of multiple sclerosis in Asahikawa, a city in northern Japan. *Journal of the Neurological Sciences*, 214 (1–2), 7–9.
- Iuliano, G., and Napoletano, R., 2008. Prevalence and incidence of multiple sclerosis in Salerno (southern Italy) and its province. *European Journal of Neurology*, 15 (1), 73–76.
- Jacobs, H. J. C., 1995. Photochemistry of conjugated trienes: Vitamin D revisited. *Pure and Applied Chemistry*, 67 (1), 63–70.
- Jacobs, H. J. C., Gielen, J. W. J., and Havinga, E., 1981. Effects of wavelength and conformation on the photochemistry of vitamin D and related conjugated trienes. *Tetrahedron Letters*, 22 (40), 4013–4016.
- Jafari, N., and Hintzen, R. Q., 2011. The association between cigarette smoking and multiple sclerosis. *Journal of the Neurological Sciences*, 311 (1–2), 78–85.
- Johnson, J. S., Nobmann, E. D., Asay, E., and Lanier, A. P., 2009. Dietary intake of Alaska Native people in two regions and implications for health: The Alaska Native dietary and subsistence food assessment project. *International Journal of Circumpolar Health*, 68 (2), 109–122.
- Jorde, R., and Bønaa, K. H., 2000. Calcium from dairy products, vitamin D intake, and blood pressure: The Tromsø study. *American Journal of Clinical Nutrition*, 71 (6), 1530–1535.
- Jurek, A. M., Greenland, S., Maldonado, G., and Church, T. R., 2005. Proper interpretation of non-differential misclassification effects: Expectations vs observations. *International Journal of Epidemiology*, 34 (3), 680–687.
- Juzeniene, A., Brekke, P., Dahlback, A., Andersson-Engels, S., Reichrath, J., Moan, K., Holick, M. F., Grant, W. B., and Moan, J., 2011. Solar radiation and human health. *Reports on Progress in Physics*, 74 (6), 066701.
- Kahana, E., 2000. Epidemiologic studies of multiple sclerosis: A review. *Biomedicine & Pharmacotherapy*, 54 (2), 100–102.
- Kakalacheva, K., and Lünemann, J. D., 2011. Environmental triggers of multiple sclerosis. *FEBS Letters*, 585 (23), 3724–3729.

- Kakalacheva, K., Münz, C., and Lünemann, J. D., 2011. Viral triggers of multiple sclerosis. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1812 (2), 132–140.
- Kampman, M., Wilsgaard, T., and Mellgren, S., 2007. Outdoor activities and diet in childhood and adolescence relate to MS risk above the Arctic circle. *Journal of Neurology*, 254, 471–477.
- Kampman, M. T., and Brustad, M., 2008. Vitamin D: A candidate for the environmental effect in multiple sclerosis — observations from Norway. *Neuroepidemiology*, 30 (3), 140–146.
- Kampman, M. T., and Steffensen, L. H., 2010. The role of vitamin D in multiple sclerosis. *Journal of Photochemistry and Photobiology B: Biology*, 101 (2), 137–141.
- Karnaikh, V. N., 2009. Эпидемиология рассеянного склероза в амурской области [Epidemiology of multiple sclerosis in Amur region]. *Zhurnal Nevrologii i Psikiatrii imeni S.S. Korsakova*, 109 (8), 59–62 (in Russian).
- Karni, A., Kahana, E., Zilber, N., Abramsky, O., Alter, M., and Karussis, D., 2003. The frequency of multiple sclerosis in Jewish and Arab populations in greater Jerusalem. *Neuroepidemiology*, 22 (1), 82–86.
- Kennedy, J., O'Connor, P., Sadovnick, A. D., Perara, M., Yee, I., and Banwell, B., 2006. Age at onset of multiple sclerosis may be influenced by place of residence during childhood rather than ancestry. *Neuroepidemiology*, 26 (3), 162–167.
- Kerr, J. B., 2005. Understanding the factors that affect surface ultraviolet radiation. *Optical Engineering*, 44 (4), 041002.
- Khanokh, E. V., Rozhdestvensky, A. S., Kudryavtseva, E. A., Kakulya, A. V., Delov, R. A., and Filipenko, M. L., 2011. Исследование наследственных факторов предрасположенности к рассеянному склерозу и особенностей его течения в русской этнической группе [Research on hereditary factors of multiple sclerosis susceptibility and peculiarities of its course in Russian ethnic group]. *Bulletin of the Siberian Branch of the Russian Academy of Medical Sciences*, 31 (1), 113–118 (in Russian).
- Khoury, S. J., Guttman, C. R. G., Orav, E. J., Kikinis, R., Jolesz, F. A., and Weiner, H. L., 2000. Changes in activated T cells in the blood correlate with disease activity in multiple sclerosis. *Archives of Neurology*, 57 (8), 1183–1189.

- Kim, H.-M., Richardson, D., Loomis, D., van Tongeren, M., and Burstyn, I., 2010a. Bias in the estimation of exposure effects with individual- or group-based exposure assessment. *Journal of Exposure Science and Environmental Epidemiology*, 21 (2), 212–221.
- Kim, N. H., Kim, H. J., Cheong, H. K., Kim, B. J., Lee, K. H., Kim, E. H., Kim, E. A., Kim, S., Park, M. S., and Yoon, W. T., 2010b. Prevalence of multiple sclerosis in Korea. *Neurology*, 75 (16), 1432–1438.
- Kimball, S. M., Ursell, M. R., O'Connor, P., and Vieth, R., 2007. Safety of vitamin D₃ in adults with multiple sclerosis. *American Journal of Clinical Nutrition*, 86 (3), 645–651.
- Kimlin, M. G., Downs, N. J., and Parisi, A. V., 2003. Comparison of human facial UV exposure at high and low latitudes and the potential impact on dermal vitamin D production. *Photochemical & Photobiological Sciences*, 2 (4), 370–375.
- Kimlin, M. G., Olds, W. J., and Moore, M. R., 2007. Location and vitamin D synthesis: Is the hypothesis validated by geophysical data? *Journal of Photochemistry and Photobiology B: Biology*, 86 (3), 234–239.
- Kioy, P. G., 2001. Emerging picture of multiple sclerosis in Kenya. *East African Medical Journal*, 78 (2), 93–96.
- Klupka-Sarić, I., Ristić, S., Sepčić, J., Kapović, M., Peterlin, B., Materljan, E., Jurišić, T., Mamić, D. M., Burina, A., and Šulentić, V., 2007. Epidemiology of multiple sclerosis in western Herzegovina. *Clinical Neurology and Neurosurgery*, 109 (9), 779–783.
- Knecht, A. K., and Bronner-Fraser, M., 2002. Induction of the neural crest: A multigene process. *Nature Reviews Genetics*, 3 (6), 453–461.
- Koch-Henriksen, N., and Sørensen, P. S., 2010. The changing demographic pattern of multiple sclerosis epidemiology. *Lancet Neurology*, 9 (5), 520–532.
- Kornek, B., and Lassmann, H., 2003. Neuropathology of multiple sclerosis—new concepts. *Brain Research Bulletin*, 61 (3), 321–326.
- Køster, B., Thorgaard, C., Philip, A., and Clemmensen, I. H., 2011. Vacations to sunny destinations, sunburn, and intention to tan: A cross-sectional study in Denmark, 2007–2009. *Scandinavian Journal of Public Health*, 39 (1), 64–69.
- Kotter, M. R., Stadelmann, C., and Hartung, H.-P., 2011. Enhancing remyelination in disease—can we wrap it up? *Brain*, DOI: 10.1093/brain/awr014.

- Krökki, O., Bloigu, R., Reunanen, M., and Remes, A., 2011. Increasing incidence of multiple sclerosis in women in Northern Finland. *Multiple Sclerosis Journal*, 17 (2), 133–138.
- Krone, B., and Grange, J., 2011. Paradigms in multiple sclerosis: Time for a change, time for a unifying concept. *Inflammopharmacology*, 19 (4), 187–195.
- Krone, B., Grange, J. M., and Goodin, D., 2011. Multiple sclerosis: Are protective immune mechanisms compromised by a complex infectious background? *Autoimmune Diseases*, 2011 (1), 101.
- Kuhlmann, T., Goldschmidt, T., Antel, J., Wegner, C., König, F., Metz, I., and Brück, W., 2009. Gender differences in the histopathology of MS? *Journal of the Neurological Sciences*, 286 (1–2), 86–91.
- Kuhnlein, H. V., Barthet, V., Farren, A., Falahi, E., Legge, D., Receveur, O., and Berti, P., 2006. Vitamins A, D, and E in Canadian Arctic traditional food and adult diets. *Journal of Food Composition and Analysis*, 19 (6–7), 495–506.
- Kuhnlein, H. V., and Receveur, O., 2007. Local cultural animal food contributes high levels of nutrients for Arctic Canadian indigenous adults and children. *Journal of Nutrition*, 137 (4), 1110–1114.
- Kurland, L. T., Stazio, A., and Reed, D., 1965. An appraisal of population studies of multiple sclerosis. *Annals of the New York Academy of Sciences*, 122 (1), 520–541.
- Kurtzke, J. F., 1977. Geography in multiple sclerosis. *Journal of Neurology*, 215 (1), 1–26.
- Kurtzke, J. F., 2000. Multiple sclerosis in time and space — geographic clues to cause. *Journal of NeuroVirology*, 6 (SUPPL. 2), S134–S140.
- Kurtzke, J. F., and Heltberg, A., 2001. Multiple sclerosis in the Faroe Islands: An epitome. *Journal of Clinical Epidemiology*, 54 (1), 1–22.
- Lagunova, Z., Porojnicu, A. C., Lindberg, F., Hexeberg, S., and Moan, J., 2009. The dependency of vitamin D status on body mass index, gender, age and season. *Anticancer Research*, 29 (9), 3713–3720.
- Landtblom, A.-M., Fazio, P., Fredrikson, S., and Granieri, E., 2010. The first case history of multiple sclerosis: Augustus d’Esté (1794–1848). *Neurological Sciences*, 31 (1), 29–33.

- Langford, I. H., Bentham, G., and McDonald, A.-L., 1998. Multi-level modelling of geographically aggregate health data: A case study on malignant melanoma mortality and UV exposure in the European community. *Statistics in Medicine*, 17, 41–57.
- Langford, M., 2006. Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. *Computers, Environment and Urban Systems*, 30 (2), 161–180.
- Larochelle, C., Alvarez, J. I., and Prat, A., 2011. How do immune cells overcome the blood–brain barrier in multiple sclerosis? *FEBS Letters*, 585 (23), 3770–3780.
- Lassmann, H., 1999. The pathology of multiple sclerosis and its evolution. *Philosophical Transactions: Biological Sciences*, 354 (1390), 1635–1640.
- Lassmann, H., 2002. Mechanisms of demyelination and tissue destruction in multiple sclerosis. *Clinical Neurology and Neurosurgery*, 104 (3), 168–171.
- Lauer, K., 2007. Sausage preservation methods and the prevalence of multiple sclerosis: An ecological study. *Ecology of Food and Nutrition*, 46 (1), 1–11.
- Lawrence, M. A., Chai, W., Kara, R., Rosenberg, I. H., Scott, J., and Tedstone, A., 2009. Examination of selected national policies towards mandatory folic acid fortification. *Nutrition Reviews*, 67, S73–S78.
- Leibowitz, U., Sharon, D., and Alter, M., 1967. Geographical considerations in multiple sclerosis. *Brain*, 90 (4), 871–886.
- Leightons. 1950. Leightons Street Map of Auckland City and Suburbs, dated ca.1950, Sir George Grey Special Collections, Auckland Libraries, NZ map 6409 Available from: Auckland City Maps Online <http://www.aucklandcity.govt.nz/> [Accessed: 2012]
- Lin, E., Armstrong-Moore, D., Liang, Z., Sweeney, J. F., Torres, W. E., Ziegler, T. R., Tangpricha, V., and Gletsu-Miller, N., 2011. Contribution of adipose tissue to plasma 25-hydroxyvitamin D concentrations during weight loss following gastric bypass surgery. *Obesity*, 19 (3), 588–594.
- Lincoln, J. A., and Cook, S. D., 2009. An overview of gene-epigenetic-environmental contributions to MS causation. *Journal of the Neurological Sciences*, 286 (1–2), 54–57.

- Link, H., Sun, J.-B., Wang, Z., Xu, Z., Löve, A., Fredrikson, S., and Olsson, T., 1992. Virus-reactive and autoreactive T cells are accumulated in cerebrospinal fluid in multiple sclerosis. *Journal of Neuroimmunology*, 38 (1–2), 63–73.
- Liu, X., and Liu, Y., 2008. The accuracy assessment in areal interpolation: An empirical investigation. *Science in China Series E: Technological Sciences*, 51 (2), 62–71.
- Liu, X. H., Kyriakidis, P. C., and Goodchild, M. F., 2008. Population-density estimation using regression and area-to-point residual kriging. *International Journal of Geographical Information Science*, 22 (4), 431–447.
- Livesey, J., Elder, P., Ellis, M. J., Florkowski, C., McKenzie, R., and Liley, B., 2007a. Response to the letter by Bolland et al on defining vitamin D deficiency. *New Zealand Medical Journal*, 120 (1265), <http://www.nzma.org.nz/journal/120-1265/2815/>.
- Livesey, J., Elder, P., Ellis, M. J., McKenzie, R., Liley, B., and Florkowski, C., 2007b. Seasonal variation in vitamin D levels in the Canterbury, New Zealand population in relation to available UV radiation. *New Zealand Medical Journal*, 120 (1262), <http://www.nzma.org.nz/journal/120-1262/2733/>.
- Lonergan, R., Kinsella, K., Fitzpatrick, P., Brady, J., Murray, B., Dunne, C., Hagan, R., Duggan, M., Jordan, S., McKenna, M., Hutchinson, M., and Tubridy, N., 2011. Multiple sclerosis prevalence in Ireland: Relationship to vitamin D status and HLA genotype. *Journal of Neurology, Neurosurgery and Psychiatry*, 82 (3), 317–322.
- Lucas, R. M., Hughes, A. M., Lay, M.-L. J., Ponsonby, A.-L., Dwyer, D. E., Taylor, B. V., and Pender, M. P., 2011a. Epstein–Barr virus and multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 82 (10), 1142–1148.
- Lucas, R. M., Ponsonby, A.-L., Dear, K., Valery, P. C., Pender, M. P., Taylor, B. V., Kilpatrick, T. J., Dwyer, T., Coulthard, A., Chapman, C., van der Mei, I., Williams, D., and McMichael, A. J., 2011b. Sun exposure and vitamin D are independent risk factors for CNS demyelination. *Neurology*, 76 (6), 540–548.
- Ludwin, S. K., and Jacobson, S., 2011. Epstein–Barr virus and MS: Causality or association? *International MS Journal*, 17 (2), 39–43.
- Lünemann, J. D., and Münz, C., 2009. EBV in MS: Guilty by association? *Trends in Immunology*, 30 (6), 243–248.

- Macdonald, H. M., Mavroei, A., Reid, D. M., and Fraser, W. D., 2010. Sunlight or diet: What is the answer for providing sufficient vitamin D in the U.K.? *British Journal of Dermatology*, 163 (2), 434–435.
- Machin, J., 2009a. xlrd: Library for developers to extract data from Microsoft Excel (tm) spreadsheet files (Version 0.7.1). Available from: <http://pypi.python.org/pypi/xlrd> [Accessed 2012].
- Machin, J., 2009b. xlwt: Library to create spreadsheet files compatible with MS Excel 97/2000/xp/2003 xls files, on any platform, with python 2.3 to 2.6 (Version 0.7.2). Available from: <http://pypi.python.org/pypi/xlwt> [Accessed 2012].
- MacLaughlin, J., and Holick, M. F., 1985. Aging decreases the capacity of human skin to produce vitamin D₃. *Journal of Clinical Investigation*, 76 (4), 1536–1538.
- MacLaughlin, J. A., Anderson, R. R., and Holick, M. F., 1982. Spectral character of sunlight modulates photosynthesis of previtamin D₃ and its photoisomers in human skin. *Science*, 216 (4549), 1001–1003.
- Maghzi, A.-H., Borazanci, A., Mcgee, J., Steven Alexander, J., Gonzalez-Toledo, E., and Minagar, A., 2011a. Multiple sclerosis: Pathophysiology, clinical features, diagnosis, and management. *In: Neuroinflammation*. London: Elsevier, 1–23.
- Maghzi, A.-H., Marta, M., Bosca, I., Savoj, M.-R., Etemadifar, M., Giovannoni, G., and Meier, U.-C., 2011b. Epstein–Barr virus and multiple sclerosis: Wrong place, wrong time? *In: Neuroinflammation*. London: Elsevier, 25–37.
- Maghzi, A. H., Ghazavi, H., Ahsan, M., Etemadifar, M., Mousavi, S. A., Khorvash, F., and Minagar, A., 2010. Increasing female preponderance of multiple sclerosis in Isfahan, Iran: A population-based study. *Multiple Sclerosis*, 16 (3), 359–361.
- Malkova, N. A., Shperling, L. P., Riabukhina, O. V., and Merkulova, E. A., 2006. Рассеянный склероз в западной сибире – двадцатипятилетнее проспективное исследование в новосибирске [Multiple sclerosis in Eastern Siberia: a 20-year prospective study in Novosibirsk city]. *Zhurnal Nevrologii i Psikiatrii imeni S.S. Korsakova, Spec No 3*, 11–16 (in Russian).
- Mallada-Frechín, J., Matías-Guiu, J., Martín, R., López-Arlandis, J. M., Camacho-Cuartero, J. M., Beltrán, I., Moltó, J. M., Falip, R., and Abellán-Mirallas, I., 2000. Prevalencia de la esclerosis múltiple en el área sanitaria de Alcoi [The prevalence of multiple sclerosis in the Alcoi Health District]. *Revista de Neurologia*, 30 (12), 1131–1134 (in Spanish).

- Mallam, E., and Scolding, N. J., 2009. The diagnosis of MS. *International MS Journal*, 16 (1), 19–25.
- Marks, R., Foley, P. A., Jolley, D., Knight, K. R., Harrison, J., and Thompson, S. C., 1995. The effect of regular sunscreen use on vitamin D levels in an Australian population: Results of a randomized controlled trial. *Archives of Dermatology*, 131 (4), 415–421.
- Marrie, R. A., 2004. Environmental risk factors in multiple sclerosis aetiology. *Lancet Neurology*, 3 (12), 709–718.
- Marrie, R. A., Yu, N., Blanchard, J., Leung, S., and Elliott, L., 2010. The rising prevalence and changing age distribution of multiple sclerosis in Manitoba. *Neurology*, 74, 465–471.
- Marrosu, M. G., Murru, R., Murru, M. R., Costa, G., Zavattari, P., Whalen, M., Cocco, E., Mancosu, C., Schirru, L., Solla, E., Fadda, E., Melis, C., Porru, I., Rolesu, M., and Cucca, F., 2001. Dissection of the HLA association with multiple sclerosis in the founder isolated population of Sardinia. *Human Molecular Genetics*, 10 (25), 2907–2916.
- Matusevicius, D., Navikas, V., Söderström, M., Xiao, B. G., Haglund, M., Fredrikson, S., and Link, H., 1996. Multiple sclerosis: The proinflammatory cytokines lymphotoxin- α and tumour necrosis factor- α are upregulated in cerebrospinal fluid mononuclear cells. *Journal of Neuroimmunology*, 66 (1–2), 115–123.
- Mayr, W. T., Pittock, S. J., McClelland, R. L., Jorgensen, N. W., Noseworthy, J. H., and Rodriguez, M., 2003. Incidence and prevalence of multiple sclerosis in Olmsted County, Minnesota, 1985–2000. *Neurology*, 61 (10), 1373–1377.
- McCarty, C. A., 2008. Sunlight exposure assessment: Can we accurately assess vitamin D exposure from sunlight questionnaires? *The American Journal of Clinical Nutrition*, 87 (4), 1097S–1101S.
- McGuigan, C., McCarthy, A., Quigley, C., Bannan, L., Hawkins, S. A., and Hutchinson, M., 2004. Latitudinal variation in the prevalence of multiple sclerosis in Ireland, an effect of genetic diversity. *Journal of Neurology, Neurosurgery and Psychiatry*, 75 (4), 572–576.
- McKenzie, R. L., Liley, J. B., and Björn, L. O., 2009. UV radiation: Balancing risks and benefits. *Photochemistry and Photobiology*, 85, 88–98.

- McLeod, J. G., Hammond, S. R., and Kurtzke, J. F., 2011. Migration and multiple sclerosis in immigrants to Australia from United Kingdom and Ireland: A reassessment. I. Risk of MS by age at immigration. *Journal of Neurology*, 258 (6), 1140–1149.
- McLintock, A. H., 1966. *An Encyclopaedia of New Zealand*, edited by A. H. McIntock, originally published in 1966. Available from: Te Ara - the Encyclopedia of New Zealand <http://www.TeAra.govt.nz/en/1966/> [Accessed: 2012]
- McLoone, P., Simics, E., Barton, A., Norval, M., and Gibbs, N. K., 2005. An action spectrum for the production of *cis*-urocanic acid in human skin *in vivo*. *Journal of Investigative Dermatology*, 124 (5), 1071–1074.
- McMichael, A. J., and Hall, A. J., 1997. Does immunosuppressive ultraviolet radiation explain the latitude gradient for multiple sclerosis? *Epidemiology*, 8 (6), 642–645.
- McQualter, J. L., and Bernard, C. C. A., 2007. Multiple sclerosis: A battle between destruction and repair. *Journal of Neurochemistry*, 100 (2), 295–306.
- Mehta, B., Ramanathan, M., and Weinstock-Guttman, B., 2011. Vitamin D and multiple sclerosis: Can vitamin D prevent disease progression? *Expert Review of Neurotherapeutics*, 11 (4), 469–471.
- Mehta, B. K., 2010. New hypotheses on sunlight and the geographic variability of multiple sclerosis prevalence. *Journal of the Neurological Sciences*, 292 (1–2), 5–10.
- Melcon, M. O., Gold, L., Carrá, A., Cáceres, F., Correale, J., Cristiano, E., Fernández Liguori, N., Garcea, O., Luetic, G., and Kremenchutzky, M., 2008. Argentine Patagonia: Prevalence and clinical features of multiple sclerosis. *Multiple Sclerosis*, 14 (5), 656–662.
- Mennis, J., and Hultgren, T., 2005. Dasymetric mapping for disaggregating coarse resolution population data.
- Merson, T. D., Binder, M. D., and Kilpatrick, T. J., 2010. Role of cytokines as mediators and regulators of microglial activity in inflammatory demyelination of the CNS. *Neuromolecular Medicine*, 12 (2), 99–132.
- Mikaeloff, Y., Caridade, G., Tardieu, M., and Suissa, S., 2007. Parental smoking at home and the risk of childhood-onset multiple sclerosis in children. *Brain*, 130 (10), 2589–2595.

- Millefiorini, E., Cortese, A., Di Rezze, S., Barletta, G., Bellantonio, P., Batocchi, A. P., Di Battista, G., Fiore, S., Gasperini, C., Grasso, M. G., Koudriatseva, T., Totaro, R., and Durastanti, V., 2010. The prevalence of multiple sclerosis in central Italy. *Multiple Sclerosis*, 16 (12), 1432–1436.
- Millen, A. E., and Bodnar, L. M., 2008. Vitamin D assessment in population-based studies: A review of the issues. *The American Journal of Clinical Nutrition*, 87 (4), 1102S–1105S.
- Miller, D. H., Hammond, S. R., McLeod, J. G., Purdie, G., and Skegg, D. C., 1990. Multiple sclerosis in Australia and New Zealand: Are the determinants genetic or environmental? *Journal of Neurology, Neurosurgery and Psychiatry*, 53 (10), 903–905.
- Moan, J., Dahlback, A., Porojnicu, A. C., and Reichrath, J., 2008. At what time should one go out in the sun? Sunlight, vitamin D and skin cancer. *Advances in Experimental Medicine and Biology*, 624, 86–88.
- Modrego, P. J., and Pina, M. A., 2003. Trends in prevalence and incidence of multiple sclerosis in Bajo Aragón, Spain. *Journal of the Neurological Sciences*, 216 (1), 89–93.
- Mohammad, T., Morrison, H., and Hogenesch, H., 1999. Urocanic acid photochemistry and photobiology. *Photochemistry and Photobiology*, 69 (2), 115–135.
- Moldovan, I. R., Cotleur, A. C., Zamor, N., Butler, R. S., and Pelfrey, C. M., 2008. Multiple sclerosis patients show sexual dimorphism in cytokine responses to myelin antigens. *Journal of Neuroimmunology*, 193 (1–2), 161–169.
- Moore, G. R. W., and Esiri, M. M., 2011. The pathology of multiple sclerosis and related disorders. *Diagnostic Histopathology*, 17 (5), 225–231.
- Moynihan, J., and Moore, H., 2010. Endocrine system dynamics and MS epidemiology. *Medical Hypotheses*, 74 (5), 814–817.
- MSIF. 2008. MS prevalence rates for WHO member states for the year 2008 [Microsoft® Office Excel spreadsheet]. Available from: <http://www.atlasofms.org/export.aspx> [Accessed: 2012].
- Mu, L., and Wang, F., 2008. A scale-space clustering method: Mitigating the effect of scale in the analysis of zone-based data. *Annals of the Association of American Geographers*, 98 (1), 85–101.

- Mugglin, A. S., Carlin, B. P., and Gelfand, A. E., 2000. Fully model-based approaches for spatially misaligned data. *Journal of the American Statistical Association*, 95 (451), 877–887.
- Munger, K. L., Levin, L. I., Hollis, B. W., Howard, N. S., and Ascherio, A., 2006. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. *Journal of the American Medical Association*, 296 (23), 2832–2838.
- Muñoz-Fernández, M. A., and Fresno, M., 1998. The role of tumour necrosis factor, interleukin 6, interferon- γ and inducible nitric oxide synthase in the development and pathology of the nervous system. *Progress in Neurobiology*, 56 (3), 307–340.
- Murray, S., Bashir, K., Penrice, G., and Womersley, S. J., 2004. Epidemiology of multiple sclerosis in Glasgow. *Scottish Medical Journal*, 49 (3), 100–104.
- Murray, T. J., 2009. The history of multiple sclerosis: The changing frame of the disease over the centuries. *Journal of the Neurological Sciences*, 277 (Suppl. 1), S3–S8.
- Murshed, S. M., 2009. Disaggregation of regional population data for residential hot water demand assessment. *In: 24th International Cartographic Conference, Santiago, Chile, Nov 15–21.*
- Navikas, V., Matusevicius, D., Söderström, M., Fredrikson, S., Kivisäkk, P., Ljungdahl, Å., Höjeberg, B., and Link, H., 1996. Increased interleukin-6 mRNA expression in blood and cerebrospinal fluid mononuclear cells in multiple sclerosis. *Journal of Neuroimmunology*, 64 (1), 63–69.
- Nicoletti, A., Lo Fermo, S., Reggio, E., Tarantello, R., Liberto, A., Le Pira, F., Patti, F., and Reggio, A., 2005. A possible spatial and temporal cluster of multiple sclerosis in the town of Linguaglossa, Sicily. *Journal of Neurology*, 252 (8), 921–925.
- Nicoletti, A., Patti, F., Lo Fermo, S., Messina, S., Bruno, E., Quattrocchi, G., Laisa, P., Cilia, S., Mostile, G., Marziolo, R., Scillieri, R., Maimone, D., and Zappia, M., 2011. Increasing frequency of multiple sclerosis in Catania, Sicily: A 30-year survey. *Multiple Sclerosis Journal*, 17 (3), 273–280.
- Nicot, A. B., 2009. Gender and sex hormones in multiple sclerosis pathology and therapy. *Frontiers in Bioscience*, 14, 4477–4515.
- Nielsen, N. M., Westergaard, T., Rostgaard, K., Frisch, M., Hjalgrim, H., Wohlfahrt, J., Koch-Henriksen, N., and Melbye, M., 2005. Familial risk of multiple sclerosis: A nationwide cohort study. *American Journal of Epidemiology*, 162 (8), 774–778.

- Nilsson, L. M., Dahlgren, L., Johansson, I., Brustad, M., Sjölander, P., and van Guelpen, B., 2011. Diet and lifestyle of the Sami of southern Lapland in the 1930s–1950s and today. *International Journal of Circumpolar Health*, 70 (3), 301–318.
- Nischwitz, S., Müller-Myhsok, B., and Weber, F., 2011. Risk conferring genes in multiple sclerosis. *FEBS Letters*, 585 (23), 3789–3797.
- NIWA. UV atlas Available from: <http://www.niwa.co.nz/our-services/online-services/uv-and-ozone/uv-atlas> [Accessed: 2012]
- Nociti, V., Frisullo, G., Marti, A., Luigetti, M., Iorio, R., Patanella, A. K., Bianco, A., Tonali, P. A., Grillo, R. L., Sabatelli, M., and Batocchi, A. P., 2010. Epstein–Barr virus antibodies in serum and cerebrospinal fluid from multiple sclerosis, chronic inflammatory demyelinating polyradiculoneuropathy and amyotrophic lateral sclerosis. *Journal of Neuroimmunology*, 225 (1–2), 149–152.
- Noonan, F. P., and De Fabo, E. C., 1992. Immunosuppression by ultraviolet B radiation: Initiation by urocanic acid. *Immunology Today*, 13 (7), 250–254.
- Norman, A. W., and Bouillon, R., 2010. Vitamin D nutritional policy needs a vision for the future. *Experimental Biology and Medicine*, 235 (9), 1034–1045.
- Norval, M., Björn, L. O., and de Gruijl, F. R., 2010. Is the action spectrum for the UV-induced production of previtamin D₃ in human skin correct? *Photochemical & Photobiological Sciences*, 9 (1), 11–17.
- Norval, M., and Wulf, H. C., 2009. Does chronic sunscreen use reduce vitamin D production to insufficient levels? *British Journal of Dermatology*, 161 (4), 732–736.
- NZDoS. 1959. *Population census, 1956*. Vol. 1–5: R. E. Owen, Govt. printer.
- NZMCH. n.d. Te Ara — the Encyclopedia of New Zealand Available from: <http://www.TeAra.govt.nz> [Accessed: 2012]
- O'Mahony, L., Stepien, M., Gibney, M. J., Nugent, A. P., and Brennan, L., 2011. The potential role of vitamin D enhanced foods in improving vitamin D status. *Nutrients*, 3 (12), 1023–1041.
- Oren, Y., Shapira, Y., Agmon-Levin, N., Kivity, S., Zafrir, Y., Altman, A., Lerner, A., and Shoenfeld, Y., 2010. Vitamin D insufficiency in a sunny environment: A demographic and seasonal analysis. *Israel Medical Association Journal*, 13 (3), 193–4.

- Orton, S.-M., Herrera, B. M., Yee, I. M., Valdar, W., Ramagopalan, S. V., Sadovnick, A. D., and Ebers, G. C., 2006. Sex ratio of multiple sclerosis in Canada: A longitudinal study. *Lancet Neurology*, 5 (11), 932–936.
- Osoegawa, M., Kira, J. I., Fukazawa, T., Fujihara, K., Kikuchi, S., Matsui, M., Kohriyama, T., Sobue, G., Yamamura, T., Itoyama, Y., Saida, T., Sakata, K., Ochi, H., Matsuoka, T., Kuroiwa, Y., Igata, A., Nishitani, H., Chiba, S., Fujii, Y., Furukawa, S., Hara, H., Hara, T., Hisanaga, K., Ikeda, S. I., Izumo, S., Kaji, R., Kanda, T., Koh, S., Kusunoki, S., Kuwabara, S., Matsuo, H., Mizusawa, H., Nakamura, T., Nomura, K., Ogino, M., Ohara, Y., Osame, M., Ota, K., Shimizu, J., Suzumura, A., Tabira, T., Tanaka, K., Tanaka, M., Yoneda, M., Yoshikawa, H., and Yuki, N., 2009. Temporal changes and geographical differences in multiple sclerosis phenotypes in Japanese: Nationwide survey results over 30 years. *Multiple Sclerosis*, 15 (2), 159–173.
- Ovesen, L., Andersen, R., and Jakobsen, J., 2003. Geographical differences in vitamin D status, with particular reference to European countries. *Proceedings of the Nutrition Society*, 62 (4), 813–821.
- Palacios, N., Alonso, A., Brønnum-Hansen, H., and Ascherio, A., 2011. Smoking and increased risk of multiple sclerosis: Parallel trends in the sex ratio reinforce the evidence. *Annals of Epidemiology*, 21 (7), 536–542.
- Palacios, N., and Simon, K. C., 2012. Smoking and multiple sclerosis, cause or coincidence? *Multiple Sclerosis and Related Disorders*, 1 (3), 106–107.
- Papathanasopoulos, P., Gourzoulidou, E., Messinis, L., Georgiou, V., and Leotsinidis, M., 2008. Prevalence and incidence of multiple sclerosis in Western Greece: A 23-year survey. *Neuroepidemiology*, 30 (3), 167–173.
- Patel, J. R., and Klein, R. S., 2011. Mediators of oligodendrocyte differentiation during remyelination. *FEBS Letters*, 585 (23), 3730–3737.
- Patsopoulos, N. A., Bayer Pharma MS Genetics Working Group, Steering Committees of Studies Evaluating IFN β -1b and a CCR1-Antagonist, ANZgene Consortium, GeneMSA, International Multiple Sclerosis Genetics Consortium, and de Bakker, P. I. W., 2011. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Annals of Neurology*, 70 (6), 897–912.

- Pekmezovic, T., Jarebinski, M., Drulovic, J., Stojisavljevic, N., and Levic, Z., 2001. Prevalence of multiple sclerosis in Belgrade, Yugoslavia. *Acta Neurologica Scandinavica*, 104 (6), 353–357.
- Pender, M., and Greer, J., 2007. Immunology of multiple sclerosis. *Current Allergy and Asthma Reports*, 7 (4), 285–292.
- Pender, M. P., 2011. The essential role of Epstein–Barr virus in the pathogenesis of multiple sclerosis. *Neuroscientist*, 17 (4), 351–367.
- Pender, M. P., Csurhes, P. A., Lenarczyk, A., Pfluger, C. M. M., and Burrows, S. R., 2009. Decreased T cell reactivity to Epstein–Barr virus infected lymphoblastoid cell lines in multiple sclerosis. *Journal of Neurology, Neurosurgery and Psychiatry*, 80 (5), 498–505.
- Perković, O., Jurjević, A., Antončić, I., Dunatov, S., Bralić, M., and Ristić, S., 2010. The town of Čabar, Croatia, familiar pseudocluster for multiple sclerosis — descriptive epidemiological study. *Collegium Antropologicum*, 34 (Suppl. 2), 141–144.
- Perron, H., Bernard, C., Bertrand, J.-B., Lang, A. B., Popa, I., Sanhadji, K., and Portoukalian, J., 2009. Endogenous retroviral genes, herpesviruses and gender in multiple sclerosis. *Journal of the Neurological Sciences*, 286 (1–2), 65–72.
- Perron, H., and Lang, A., 2010. The human endogenous retrovirus link between genes and environment in multiple sclerosis and in multifactorial diseases associating neuroinflammation. *Clinical Reviews in Allergy and Immunology*, 39 (1), 51–61.
- Peterlin, B., Ristić, S., Sepčić, J., Vračko, B. K., Rako, A., Lovrečić, L., Brajenović-Milić, B., Rudež, J., Materljan, E., and Kapović, M., 2006. Region with persistent high frequency of multiple sclerosis in Croatia and Slovenia. *Journal of the Neurological Sciences*, 247 (2), 169–172.
- Petermann, F., and Korn, T., 2011. Cytokines and effector T cell subsets causing autoimmune CNS disease. *FEBS Letters*, 585 (23), 3747–3757.
- Petrovskii, S., and Li, B.-L., 2003. An exactly solvable model of population dynamics with density-dependent migrations and the allee effect. *Mathematical Biosciences*, 186 (1), 79–91.
- Phillips, K. M., Ruggio, D. M., Horst, R. L., Minor, B., Simon, R. R., Feeney, M. J., Byrdwell, W. C., and Haytowitz, D. B., 2011. Vitamin D and sterol composition

- of 10 types of mushrooms from retail suppliers in the United States. *Journal of Agricultural and Food Chemistry*, 59 (14), 7841–7853.
- Pierrot-Deseilligny, C., and Souberbielle, J. C., 2010. Is hypovitaminosis D one of the environmental risk factors for multiple sclerosis? *Brain*, 133 (7), 1869–1888.
- Pohl, D., 2009. Epstein–Barr virus and multiple sclerosis. *Journal of the Neurological Sciences*, 286 (1–2), 62–64.
- Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F. D., Montalban, X., O'Connor, P., Sandberg-Wollheim, M., Thompson, A. J., Waubant, E., Weinshenker, B., and Wolinsky, J. S., 2011. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of Neurology*, 69 (2), 292–302.
- Ponsonby, A.-L., Lucas, R. M., Lewis, S., and Halliday, J., 2010. Vitamin D status during pregnancy and aspects of offspring health. *Nutrients*, 2 (3), 389–407.
- Ponsonby, A.-L., McMichael, A., and van der Mei, I., 2002. Ultraviolet radiation and autoimmune disease: Insights from epidemiological research. *Toxicology*, 181–182, 71–78.
- Ponsonby, A. L., Lucas, R. M., and van der Mei, I. A. F., 2005. UVR, vitamin D and three autoimmune diseases—multiple sclerosis, type 1 diabetes, rheumatoid arthritis. *Photochemistry and Photobiology*, 81, 1267–1275.
- Pope, S. J., Holick, M. F., Mackin, S., and Godar, D. E., 2008. Action spectrum conversion factors that change erythemally weighted to previtamin D₃-weighted UV doses. *Photochemistry and Photobiology*, 84, 1277–1283.
- Pugliatti, M., Cossu, P., Sotgiu, S., Rosati, G., and Riise, T., 2009. Clustering of multiple sclerosis, age of onset and gender in Sardinia. *Journal of the Neurological Sciences*, 286 (1–2), 6–13.
- Pugliatti, M., Rosati, G., Carton, H., Riise, T., Drulovic, J., Vécsei, L., and Milanov, I., 2006. The epidemiology of multiple sclerosis in Europe. *European Journal of Neurology*, 13 (7), 700–722.
- Pugliatti, M., Sotgiu, S., and Rosati, G., 2002. The worldwide prevalence of multiple sclerosis. *Clinical Neurology and Neurosurgery*, 104, 182–191.
- Pugliatti, M., Sotgiu, S., Solinas, G., Castiglia, P., Pirastu, M. I., Murgia, B., Mannu, L., Sanna, G., and Rosati, G., 2001a. Multiple sclerosis epidemiology in Sardinia:

- Evidence for a true increasing risk. *Acta Neurologica Scandinavica*, 103 (1), 20–26.
- Pugliatti, M., Sotgiu, S., Solinas, G., Castiglia, P., and Rosati, G., 2001b. Multiple sclerosis prevalence among Sardinians: Further evidence against the latitude gradient theory. *Neurological Sciences*, 22 (2), 163–165.
- pveducation. PVCDROM Available from: <http://www.pveducation.org> [Accessed: 2012]
- R Development Core Team. 2009. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: <http://www.R-project.org> [Accessed 2012].
- Ramagopalan, S. V., Dobson, R., Meier, U. C., and Giovannoni, G., 2010. Multiple sclerosis: Risk factors, prodromes, and potential causal pathways. *Lancet Neurology*, 9 (7), 727–739.
- Ramagopalan, S. V., Hoang, U., Seagroatt, V., Handel, A., Ebers, G. C., Giovannoni, G., and Goldacre, M. J., 2011. Geography of hospital admissions for multiple sclerosis in England and comparison with the geography of hospital admissions for infectious mononucleosis: A descriptive study. *Journal of Neurology, Neurosurgery and Psychiatry*, 82 (6), 682–687.
- Ramagopalan, S. V., Maugeri, N. J., Handunnetthi, L., Lincoln, M. R., Orton, S.-M., Dyment, D. A., Deluca, G. C., Herrera, B. M., Chao, M. J., Sadovnick, A. D., Ebers, G. C., and Knight, J. C., 2009. Expression of the multiple sclerosis-associated MHC Class II allele *HLA-DRB1*1501* is regulated by vitamin D. *PLoS Genetics*, 5 (2), e1000369.
- Rauchway, A. C., 2011. Neutralizing antibodies and multiple sclerosis. *In: Neuroinflammation*. London: Elsevier, 39–53.
- Reibel, M., 2007. Geographic information systems and spatial data processing in demography: A review. *Population Research and Policy Review*, 26 (5), 601–618.
- Rhodes, L. E., Webb, A. R., Fraser, H. I., Kift, R., Durkin, M. T., Allan, D., O'Brien, S. J., Vail, A., and Berry, J. L., 2010. Recommended summer sunlight exposure levels can produce sufficient (≥ 20 ng ml⁻¹) but not the proposed optimal (≥ 32 ng ml⁻¹) 25(OH)D levels at UK latitudes. *Journal of Investigative Dermatology*, 130 (5), 1411–1418.

- Riise, T., Nortvedt, M. W., and Ascherio, A., 2003. Smoking is a risk factor for multiple sclerosis. *Neurology*, 61 (8), 1122–1124.
- Rioux, J. D., and Abbas, A. K., 2005. Paths to understanding the genetic basis of autoimmune disease. *Nature*, 435 (7042), 584–589.
- Risco, J., Maldonado, H., Luna, L., Osada, J., Ruiz, P., Juarez, A., and Vizcarra, D., 2011. Latitudinal prevalence gradient of multiple sclerosis in Latin America. *Multiple Sclerosis Journal*, 17 (9), 1055–1059.
- Robertson, N., and Compston, A., 1995. Surveying multiple sclerosis in the United Kingdom. *Journal of Neurology, Neurosurgery and Psychiatry*, 58 (1), 2–6.
- Robinson, A., Becker, R., the ReportLab team, and the community. 2010. reportlab: The ReportLab toolkit (Version 2.4). Available from: <http://www.reportlab.com/> [Accessed 2012].
- Robson, B., Purdie, G., Cram, F., and Simmonds, S., 2007. Age standardisation — an indigenous standard? *Emerging Themes in Epidemiology*, 4 (1), 3.
- Rockell, J., Skeaff, C. M., Logan, V., and Green, T., 2008. Vitamin D: A review prepared for the New Zealand Food Safety Authority and the Ministry of Health. *New Zealand Food Safety Authority*. Available from: http://www.foodsafety.govt.nz/elibrary/industry/vitamin-review-prepared-research-projects/Vitamin_D_Final_Report_Otago_University.pdf [Accessed: 2012].
- Rockell, J. E. P., Skeaff, C. M., Williams, S. M., and Green, T. J., 2006. Serum 25-hydroxyvitamin D concentrations of New Zealanders aged 15 years and older. *Osteoporosis International*, 17 (9), 1382–9.
- Rodriguez, M., 2009. Have we finally identified an autoimmune demyelinating disease? *Annals of Neurology*, 66 (5), 572–573.
- Rosati, G., 1994. Descriptive epidemiology of multiple sclerosis in Europe in the 1980s: A critical overview. *Annals of Neurology*, 36 (S2), S164–S174.
- Rosati, G., 2001. The prevalence of multiple sclerosis in the world: An update. *Neurological Sciences*, 22 (2), 117–139.
- Rothman, K. J., 2002. *Epidemiology : An introduction*. Oxford University Press, Inc.
- Rovaris, M., Confavreux, C., Furlan, R., Kappos, L., Comi, G., and Filippi, M., 2006. Secondary progressive multiple sclerosis: Current knowledge and future challenges. *Lancet Neurology*, 5 (4), 343–354.

- Rovner, A. J., and O'Brien, K. O., 2008. Hypovitaminosis D among healthy children in the United States: A review of the current evidence. *Archives of Pediatrics & Adolescent Medicine*, 162 (6), 513–519.
- Rushton, W. A. H., 1951. A theory of the effects of fibre size in medullated nerve. *Journal of Physiology*, 115 (1), 101–122.
- Saadatnia, M., Etemadifar, M., and Maghzi, A. H., 2007. Multiple sclerosis in Isfahan, Iran. *International Review of Neurobiology*, 79, 357–375.
- Sabel, C. E., Gatrell, A. C., Löytönen, M., Maasilta, P., and Jokelainen, M., 2000. Modelling exposure opportunities: Estimating relative risk for motor neurone disease in Finland. *Social Science & Medicine*, 50 (7-8), 1121–1137.
- Sadahiro, Y., 1999. Accuracy of areal interpolation: A comparison of alternative methods. *Journal of Geographical Systems*, 323–346.
- Sadahiro, Y., 2000. Accuracy of count data transferred through the areal weighting interpolation method. *International Journal of Geographical Information Science*, 14 (1), 25–50.
- Sagnes-Raffy, C., Gourraud, P. A., Hannon, V., Bourrel, R., Laffontan, M. A., Gaulene, M. C., Viala, F., and Clanet, M., 2010. La SEP en Haute-Garonne: Une sous-estimation importante du nombre de cas [Multiple sclerosis in Haute-Garonne: An important underestimate of prevalence]. *Revue d'Épidémiologie et de Santé Publique*, 58 (1), 23–31 (in French).
- Salemi, G., Ragonese, P., Aridon, P., Scola, G., Saporito, V., Conte, S., and Savettieri, G., 2000. Incidence of multiple sclerosis in Bagheria city, Sicily, Italy. *Neurological Sciences*, 21 (6), 361–365.
- Salzer, J., Svenningsson, A., and Sundström, P., 2010. Season of birth and multiple sclerosis in Sweden. *Acta Neurologica Scandinavica*, 121 (1), 20–23.
- Sánchez, J. L., Aguirre, C., Arcos-Burgos, O. M., Jiménez, I., Jiménez, M., León, F., Pareja, J., Pradilla, G., Uribe, B., Uribe, C. S., Villa, A., Volcy, M., and Palacio, L. G., 2000. Prevalencia de la esclerosis múltiple en Colombia [Prevalence of multiple sclerosis in Colombia]. *Revista de Neurología*, 31 (12), 1101–1103 (in Spanish).
- Saporito, S., Chavers, J. M., Nixon, L. C., and McQuiddy, M. R., 2007. From here to there: Methods of allocating data between census geography and socially meaningful areas. *Social Science Research*, 36 (3), 897–920.

- Sarasoja, T., Wikström, J., Paltamaa, J., Hakama, M., and Sumelahti, M. L., 2004. Occurrence of multiple sclerosis in central Finland: A regional and temporal comparison during 30 years. *Acta Neurologica Scandinavica*, 110 (5), 331–336.
- Sato, F., Omura, S., Martinez, N. E., and Tsunoda, I., 2011. Animal models of multiple sclerosis. *In: Neuroinflammation*. London: Elsevier, 55–79.
- Sawcer, S., and Hellenthal, G., 2011. The major histocompatibility complex and multiple sclerosis: A smoking gun? *Brain*, 134 (3), 638–640.
- Scharla, S. H., 1998. Prevalence of subclinical vitamin D deficiency in different European countries. *Osteoporosis International, Suppl. 8*, S7–12.
- Schirmer, L., Antel, J. P., Brück, W., and Stadelmann, C., 2011. Axonal loss and neurofilament phosphorylation changes accompany lesion development and clinical progression in multiple sclerosis. *Brain Pathology*, 21, 428–440.
- Schmidt, H., Williamson, D., and Ashley-Koch, A., 2007. HLA-DR15 haplotype and multiple sclerosis: A HuGE review. *American Journal of Epidemiology*, 165 (10), 1097–1109.
- Scragg, R., and Bartley, J., 2007. Vitamin D—how do we define deficiency and what can we do about it in New Zealand? *New Zealand Medical Journal*, 120 (1262), <http://www.nzma.org.nz/journal/120-1262/2735/>.
- Segi, M., 1960. *Cancer mortality for selected sites in 24 countries (1950–57)*. Department of Public Health, Tohoku University of Medicine, Sendai, Japan.
- Sellner, J., Kraus, J., Awad, A., Milo, R., Hemmer, B., and Stüve, O., 2011. The increasing incidence and prevalence of female multiple sclerosis—a critical analysis of potential environmental factors. *Autoimmunity Reviews*, 10 (8), 495–502.
- Shapira, Y., Agmon-Levin, N., and Shoenfeld, Y., 2010. Defining and analyzing geoepidemiology and human autoimmunity. *Journal of Autoimmunity*, 34 (3), J168–J177.
- Sharma, S., Barr, A. B., Macdonald, H. M., Sheehy, T., Novotny, R., and Corriveau, A., 2011. Vitamin D deficiency and disease risk among aboriginal Arctic populations. *Nutrition Reviews*, 69 (8), 468–478.
- Shroff, R., Knott, C., and Rees, L., 2010. The virtues of vitamin D—but how much is too much? *Pediatric Nephrology*, 25 (9), 1607–1620.

- Simmons, R. D., Hall, C. A., Gleeson, P., Everard, G., Casse, R. F., and O'Brien, E. D., 2001. Prevalence survey of multiple sclerosis in the Australian Capital Territory. *Internal Medicine Journal*, 31 (3), 161–167.
- Simon, K. C., van der Mei, I. A. F., Munger, K. L., Ponsonby, A., Dickinson, J., Dwyer, T., Sundström, P., and Ascherio, A., 2010. Combined effects of smoking, anti-EBNA antibodies, and HLA-DRB1*1501 on multiple sclerosis risk. *Neurology*, 74 (17), 1365–1371.
- Simpson, S., Blizzard, L., Otahal, P., van der Mei, I., and Taylor, B., 2011. Latitude is significantly associated with the prevalence of multiple sclerosis: A meta-analysis. *Journal of Neurology, Neurosurgery and Psychiatry*, 82 (10), 1132–1141.
- Sinclair, C., and Foley, P., 2009. Skin cancer prevention in Australia. *British Journal of Dermatology*, 161 (Suppl. 3), 116–123.
- Sivertseva, S. A., Kandala, N. S., Zhuravlev, M. N., Zakatei, I. G., and Boïko, A. N., 2010. Рассеянный склероз у коренных жителей ямала [Multiple sclerosis in the native population of Yamal]. *Zhurnal Nevrologii i Psikhatrii imeni S.S. Korsakova*, 110 (1), 97–99 (in Russian).
- Sivertseva, S. A., Zhuravlev, M. N., Murav'ev, S. A., and Boïko, A. N., 2006. Эпидемиология рассеянного склероза в тюменской области [Epidemiology of multiple sclerosis in Tyumen region]. *Zhurnal Nevrologii i Psikhatrii imeni S.S. Korsakova, Spec No 3*, 22–25 (in Russian).
- Skulina, C., Schmidt, S., Dornmair, K., Babbe, H., Roers, A., Rajewsky, K., Wekerle, H., Hohlfeld, R., and Goebels, N., 2004. Multiple sclerosis: Brain-infiltrating CD8+ T cells persist as clonal expansions in the cerebrospinal fluid and blood. *Proceedings of the National Academy of Sciences*, 101 (8), 2428–2433.
- Sloka, J. S., Pryse-Phillips, W. E. M., and Stefanelli, M., 2005a. Incidence and prevalence of multiple sclerosis in Newfoundland and Labrador. *Canadian Journal of Neurological Sciences*, 32 (1), 37–42.
- Sloka, J. S., Pryse-Phillips, W. E. M., and Stefanelli, M., 2005b. Multiple sclerosis in Newfoundland and Labrador — a model for disease prevalence. *Canadian Journal of Neurological Sciences*, 32 (1), 43–49.

- Sloka, S., Silva, C., Pryse-Phillips, W., Patten, S., Metz, L., and Yong, V. W., 2011. A quantitative analysis of suspected environmental causes of MS. *Canadian Journal of Neurological Sciences*, 38 (1), 98–105.
- Slominski, A., and Wortsman, J., 2000. Neuroendocrinology of the skin. *Endocrine Reviews*, 21 (5), 457–487.
- Smestad, C., Sandvik, L., Holmoy, T., Harbo, H. F., and Celius, E. G., 2008. Marked differences in prevalence of multiple sclerosis between ethnic groups in Oslo, Norway. *Journal of Neurology*, 255 (1), 49–55.
- Smith, S. M., Gardner, K. K., Locke, J., and Zwart, S. R., 2009. Vitamin D supplementation during Antarctic winter. *American Journal of Clinical Nutrition*, 89 (4), 1092–1098.
- Smolders, J., Damoiseaux, J., Menheere, P., and Hupperts, R., 2008. Vitamin D as an immune modulator in multiple sclerosis, a review. *Journal of Neuroimmunology*, 194 (1–2), 7–17.
- Smolders, J., Moen, S. M., Damoiseaux, J., Huitinga, I., and Holmøy, T., 2011. Vitamin D in the healthy and inflamed central nervous system: Access and function. *Journal of the Neurological Sciences*, 311 (1–2), 37–43.
- Solaro, C., Allemani, C., Messmer Uccelli, M., Canevari, E., Dagnino, N., Pizio, R., Regesta, G., Tanganelli, P., Battaglia, M. A., and Mancardi, G. L., 2005. The prevalence of multiple sclerosis in the north-west Italian province of Genoa. *Journal of Neurology*, 252 (4), 436–440.
- Solomon, A., and Whitham, R., 2010. Multiple sclerosis and vitamin D: A review and recommendations. *Current Neurology and Neuroscience Reports*, 10 (5), 389–396.
- Sosa, R. A., and Forsthuber, T. G., 2011. The critical role of antigen-presentation-induced cytokine crosstalk in the central nervous system in multiple sclerosis and experimental autoimmune encephalomyelitis. *Journal of Interferon and Cytokine Research*, 31 (10), 753–768.
- Sospedra, M., and Martin, R., 2005. Immunology of multiple sclerosis. *Annual Review of Immunology*, 23, 683–747.
- Sotgiu, S., Pugliatti, M., Fois, M. L., Arru, G., Sanna, A., Sotgiu, M. A., and Rosati, G., 2004. Genes, environment, and susceptibility to multiple sclerosis. *Neurobiology of Disease*, 17 (2), 131–143.

- Springbett, P., Buglass, S., and Young, A. R., 2010. Photoprotection and vitamin D status. *Journal of Photochemistry and Photobiology B: Biology*, 101 (2), 160–168.
- Spurkland, A., Celius, E. G., Knutsen, I., Belske, A., Thorsby, E., and Vartdal, F., 1997. The HLA-DQ($\alpha 1^*0102$, $\beta 1^*0602$) heterodimer may confer susceptibility to multiple sclerosis in the absence of the HLA-DR($\alpha 1^*01$, $\beta 1^*1501$) heterodimer. *Tissue Antigens*, 50 (1), 15–22.
- Sriram, S., 2011. Role of glial cells in innate immunity and their role in CNS demyelination. *Journal of Neuroimmunology*, 239 (1–2), 13–20.
- Stangel, M., Pul, R., Skripuletz, T., Trebst, C., Voss, E., and Gudi, V., 2011. Remyelination in multiple sclerosis. *In: Neuroinflammation*. London: Elsevier, 193–223.
- Staples, J., Ponsonby, A.-L., and Lim, L., 2010. Low maternal exposure to ultraviolet radiation in pregnancy, month of birth, and risk of multiple sclerosis in offspring: Longitudinal analysis. *BMJ*, 340, c1640.
- Statistics Norway. 2010. Sami statistics 2010 *Official Statistics of Norway* Available from:
http://www.ssb.no/english/subjects/00/00/20/nos_d443_en/nos_d443_en.pdf
 [Accessed: 2012]
- StatsNZ. n.d. Stats NZ Available from: <http://www.stats.govt.nz/> [Accessed: 2012]
- Stefano, S., Maura, P., Alessandra, S., and Giulio, R., 2003. Does the "hygiene hypothesis" provide an explanation for the high prevalence of multiple sclerosis in Sardinia? *Autoimmunity*, 36 (5), 257–260.
- Stenager, E., Møller, R., Brønnum-Hansen, H., and Koch-Henriksen, N. J., 2005. Multipel sklerose: Udvikling i prævalens i Danmark 1950–1996 samt prævalens i amterne 1996 [Multiple sclerosis: Its change in prevalence in Denmark, 1950–1996, and its total prevalence in Danish counties in 1996]. *Ugeskrift for Læger*, 167 (14), 1500–1502 (in Danish).
- Su, M.-D., Lin, M.-C., Hsieh, H.-I., Tsai, B.-W., and Lin, C.-H., 2010. Multi-layer multi-class dasymetric mapping to estimate population distribution. *Science of The Total Environment*, 408 (20), 4807–4816.

- Sumelahti, M. L., Tienari, P. J., Wikström, J., Palo, J., and Hakama, M., 2001. Increasing prevalence of multiple sclerosis in Finland. *Acta Neurologica Scandinavica*, 103 (3), 153–158.
- Sundström, P., Nyström, L., and Forsgren, L., 2001. Prevalence of multiple sclerosis in Västerbotten county in northern Sweden. *Acta Neurologica Scandinavica*, 103 (4), 214–218.
- Sundström, P., Nyström, L., and Forsgren, L., 2003. Incidence (1988-97) and prevalence (1997) of multiple sclerosis in Västerbotten county in northern Sweden. *Journal of Neurology, Neurosurgery and Psychiatry*, 74 (1), 29–32.
- Sundström, P., Nyström, M., Ruuth, K., and Lundgren, E., 2009. Antibodies to specific EBNA-1 domains and HLA DRB1*1501 interact as risk factors for multiple sclerosis. *Journal of Neuroimmunology*, 215 (1–2), 102–107.
- Svenson, L. W., Warren, S., Warren, K. G., Metz, L. M., Patten, S. B., and Schopflocher, D. P., 2007. Prevalence of multiple sclerosis in First Nations people of Alberta. *Canadian Journal of Neurological Sciences*, 34 (2), 175–180.
- Swingler, R. J., and Compston, D., 1986. The distribution of multiple sclerosis in the United Kingdom. *Journal of Neurology, Neurosurgery and Psychiatry*, 49 (10), 1115–1124.
- Tatalovich, Z., Wilson, J. P., Mack, T., Yan, Y., and Cockburn, M., 2006. The objective assessment of lifetime cumulative ultraviolet exposure for determining melanoma risk. *Journal of Photochemistry and Photobiology B: Biology*, 85 (3), 198–204.
- Taylor, B. V., Lucas, R. M., Dear, K., Kilpatrick, T. J., Pender, M. P., van der Mei, I. A. F., Chapman, C., Coulthard, A., Dwyer, T., McMichael, A. J., Valery, P. C., Williams, D., and Ponsonby, A.-L., 2010a. Latitudinal variation in incidence and type of first central nervous system demyelinating events. *Multiple Sclerosis*, 16 (4), 398–405.
- Taylor, B. V., Pearson, J. F., Clarke, G., Mason, D. F., Abernethy, D. A., Willoughby, E., and Sabel, C., 2010b. MS prevalence in New Zealand, an ethnically and latitudinally diverse country. *Multiple Sclerosis*, 16 (12), 1422–1431.
- Thacker, E. L., Mirzaei, F., and Ascherio, A., 2006. Infectious mononucleosis and risk for multiple sclerosis: A meta-analysis. *Annals of Neurology*, 59, 499–503.

- Thomas, D., Stram, D., and Dwyer, J., 1993. Exposure measurement error: Influence on exposure–disease relationships and methods of correction. *Annual Review of Public Health*, 14, 69–93.
- Thompson, A. J., Polman, C. H., Miller, D. H., McDonald, W. I., Brochet, B., Filippi, M., Montalban, X., and De Sá, J., 1997. Primary progressive multiple sclerosis. *Brain*, 120 (6), 1085–1096.
- Thompson, R. L., 2007. Is low vitamin D status a problem in Britain? *Nutrition Bulletin*, 32 (4), 311–313.
- Tiebei, L., David, P., Jonathan, C., and Robert, S., 2007. A comparison of spatial disaggregation techniques as applied to population estimation for South East Queensland (SEQ), Australia. *Applied GIS*, 3 (9), 1–16.
- Tomassini, V., and Pozzilli, C., 2009. Sex hormones, brain damage and clinical course of multiple sclerosis. *Journal of the Neurological Sciences*, 286 (1–2), 35–39.
- Toro, J., Sarmiento, O. L., Díaz Del Castillo, A., Satizábal, C. L., Ramírez, J. D., Montenegro, A. C., Góngora, M. C., Quiñones, J. A., Díaz, A., and Tobón, A., 2007. Prevalence of multiple sclerosis in Bogotá, Colombia. *Neuroepidemiology*, 28 (1), 33–38.
- Traugott, U., and Lebon, P., 1988. Interferon- γ and Ia antigen are present on astrocytes in active chronic multiple sclerosis lesions. *Journal of the Neurological Sciences*, 84 (2–3), 257–264.
- Tsai, C. P., Yuan, C. L., Yu, H. Y., Chen, C., Guo, Y. C., and Shan, D. E., 2004. Multiple sclerosis in Taiwan. *Journal of the Chinese Medical Association*, 67 (10), 500–505.
- Tsiaras, W. G., and Weinstock, M. A., 2011. Factors influencing vitamin D status. *Acta Dermato-Venereologica*, 91 (2), 115–124.
- Tumani, H., Hartung, H.-P., Hemmer, B., Teunissen, C., Deisenhammer, F., Giovannoni, G., and Zettl, U. K., 2009. Cerebrospinal fluid biomarkers in multiple sclerosis. *Neurobiology of Disease*, 35 (2), 117–127.
- Turabelidze, G., Schootman, M., Zhu, B. P., Malone, J. L., Horowitz, S., Weidinger, J., Williamson, D., and Simoes, E., 2008. Multiple sclerosis prevalence and possible lead exposure. *Journal of the Neurological Sciences*, 269 (1-2), 158–162.
- Tylavsky, F. A., Cheng, S., Lyytikäinen, A., Viljakainen, H., and Lamberg-Allardt, C., 2006. Strategies to improve vitamin D status in northern European children: Exploring

- the merits of vitamin D fortification and supplementation. *Journal of Nutrition*, 136 (4), 1130–1134.
- Unger, M. D., Cuppari, L., Titan, S. M., Magalhães, M. C. T., Sasaki, A. L., Dos Reis, L. M., Jorgetti, V., and Moysés, R. M. A., 2010. Vitamin D status in a sunny country: Where has the sun gone? *Clinical Nutrition*, 29 (6), 784–788.
- Van Belle, T. L., Gysemans, C., and Mathieu, C., 2011. Vitamin D in autoimmune, infectious and allergic diseases: A vital player? *Best Practice & Research Clinical Endocrinology & Metabolism*, 25 (4), 617–632.
- van der Mei, I. A. F., Ponsonby, A.-L., Dwyer, T., Blizzard, L., Simmons, R., Taylor, B. V., Butzkueven, H., and Kilpatrick, T., 2003. Past exposure to sun, skin phenotype, and risk of multiple sclerosis: Case-control study. *BMJ*, 327 (7410), 316.
- van der Mei, I. A. F., Ponsonby, A.-L., Dwyer, T., Blizzard, L., Taylor, B. V., Kilpatrick, T., Butzkueven, H., and McMichael, A. J., 2007a. Vitamin D levels in people with multiple sclerosis and community controls in Tasmania, Australia. *Journal of Neurology*, 254 (0340-5354 (Print)), 581–590.
- van der Mei, I. A. F., Ponsonby, A.-L., Engelsen, O., Pasco, J. A., McGrath, J. J., Eyles, D. W., Blizzard, L., Dwyer, T., Lucas, R., and Jones, G., 2007b. The high prevalence of vitamin D insufficiency across Australian populations is only partly explained by season and latitude. *Environmental Health Perspectives*, 115 (8), 1132–1139.
- van Rossum, G., and PythonLabs Team. 2007. Python (Version 2.5.1): Python Software Foundation. Available from: <http://www.python.org> [Accessed 2012].
- Vieth, R., 2001. Vitamin D nutrition and its potential health benefits for bone, cancer and other conditions. *Journal of Nutritional and Environmental Medicine*, 11 (4), 275–291.
- Vieth, R., 2005. The pharmacology of vitamin D, including fortification strategies. *In*: Feldman, D., Pike, J. W., and Glorieux, F., H. eds. *Vitamin D* 2nd ed. Burlington: Academic Press, 995–1015.
- Vieth, R., 2006. Critique of the considerations for establishing the tolerable upper intake level for vitamin D: Critical need for revision upwards. *Journal of Nutrition*, 136 (4), 1117–1122.

- Visscher, B. R., Detels, R., Coulson, A. H., Malmgren, R. M., and Dudley, J. P., 1977. Latitude, migration, and the prevalence of multiple sclerosis. *American Journal of Epidemiology*, 106 (6), 470–475.
- Vukusic, S., Van Bockstael, V., Gosselin, S., and Confavreux, C., 2007. Regional variations in the prevalence of multiple sclerosis in French farmers. *Journal of Neurology, Neurosurgery and Psychiatry*, 78 (7), 707–709.
- Wallin, M. T., and Kurtzke, J. F., 2011a. Comments on familial multiple sclerosis in the Faroe Islands: Letter to the editor. *Acta Neurologica Scandinavica*, 121 (6), 429–431.
- Wallin, M. T., and Kurtzke, J. F., 2011b. Multiple sclerosis in the Faroe Islands: A core context for neuroanthropology. *Acta Neurologica Scandinavica*, 124 (1), 66–67.
- Wallin, M. T., Page, W. F., and Kurtzke, J. F., 2004. Multiple sclerosis in US veterans of the Vietnam era and later military service: Race, sex, and geography. *Annals of Neurology*, 55 (1), 65–71.
- Wallin, M. T., Page, W. F., and Kurtzke, J. F., 2009. Migration and multiple sclerosis in Alaskan military veterans. *Journal of Neurology*, 256 (9), 1413–1417.
- Wandstrat, A., and Wakeland, E., 2001. The genetics of complex autoimmune diseases: Non-MHC susceptibility genes. *Nature immunology*, 2 (9), 802–809.
- Wang, T. J., Zhang, F., Richards, J. B., Kestenbaum, B., Van Meurs, J. B., Berry, D., Kiel, D. P., Streeten, E. A., Ohlsson, C., Koller, D. L., Peltonen, L., Cooper, J. D., O'Reilly, P. F., Houston, D. K., Glazer, N. L., Vandenput, L., Peacock, M., Shi, J., Rivadeneira, F., McCarthy, M. I., Anneli, P., de Boer, I. H., Mangino, M., Kato, B., Smyth, D. J., Booth, S. L., Jacques, P. F., Burke, G. L., Goodarzi, M., Cheung, C.-L., Wolf, M., Rice, K., Goltzman, D., Hidiroglou, N., Ladouceur, M., Wareham, N. J., Hocking, L. J., Hart, D., Arden, N. K., Cooper, C., Malik, S., Fraser, W. D., Hartikainen, A.-L., Zhai, G., Macdonald, H. M., Forouhi, N. G., Loos, R. J. F., Reid, D. M., Hakim, A., Dennison, E., Liu, Y., Power, C., Stevens, H. E., Jaana, L., Vasan, R. S., Soranzo, N., Bojunga, J., Psaty, B. M., Lorentzon, M., Foroud, T., Harris, T. B., Hofman, A., Jansson, J.-O., Cauley, J. A., Uitterlinden, A. G., Gibson, Q., Järvelin, M.-R., Karasik, D., Siscovick, D. S., Econs, M. J., Kritchevsky, S. B., Florez, J. C., Todd, J. A., Dupuis, J., Hyppönen, E., and Spector, T. D., 2010. Common genetic determinants of vitamin D insufficiency: A genome-wide association study. *Lancet*, 376 (9736), 180–188.

- Warren, S., Warren, K. G., and Turpin, K. V. L., 2011. Ultraviolet radiation (UVR) and risk of developing multiple sclerosis (MS). *Canadian Journal of Neurological Sciences*, 38 (1), 6–7.
- Warren, S. A., Svenson, L. W., and Warren, K. G., 2008. Contribution of incidence to increasing prevalence of multiple sclerosis in Alberta, Canada. *Multiple Sclerosis*, 14 (7), 872–879.
- Webb, A. R., Decosta, B. R., and Holick, M. F., 1989. Sunlight regulates the cutaneous production of vitamin D₃ by causing its photodegradation. *Journal of Clinical Endocrinology and Metabolism*, 68 (5), 882–887.
- Webb, A. R., and Engelsen, O., 2006. Calculated ultraviolet exposure levels for a healthy vitamin D status. *Photochemistry and Photobiology*, 82, 1697–1703.
- Webb, A. R., and Holick, M. F., 1988. The role of sunlight in the cutaneous production of vitamin D₃. *Annual Review of Nutrition*, 8, 375–399.
- Webb, A. R., Kift, R., Berry, J. L., and Rhodes, L. E., 2011a. The vitamin D debate: Translating controlled experiments into reality for human sun exposure times. *Photochemistry and Photobiology*, 87 (3), 741–745.
- Webb, A. R., Kift, R., Durkin, M. T., O'Brien, S. J., Vail, A., Berry, J. L., and Rhodes, L. E., 2010. The role of sunlight exposure in determining the vitamin D status of the U.K. white adult population. *British Journal of Dermatology*, 163 (5), 1050–1055.
- Webb, A. R., Kline, L., and Holick, M. F., 1988. Influence of season and latitude on the cutaneous synthesis of vitamin D₃: Exposure to winter sunlight in Boston and Edmonton will not promote vitamin D₃ synthesis in human skin. *Journal of Clinical Endocrinology & Metabolism*, 67 (2), 373–378.
- Webb, A. R., Slaper, H., Koepke, P., and Schmalwieser, A. W., 2011b. Know your standard: Clarifying the CIE erythema action spectrum. *Photochemistry and Photobiology*, 87, 483–486.
- Webster, T. F., 2007. Bias magnification in ecologic studies: A methodological investigation. *Environmental Health*, 6, doi:10.1186/1476-069X-6-17.
- Welch, A. A., Lund, E., Amiano, P., Dorronsoro, M., Brustad, M., Kumle, M., Rodriguez, M., Lasheras, C., Janzon, L., Jansson, J., Luben, R., Spencer, E. A., Overvad, K., Tjønneland, A., Clavel-Chapelon, F., Linseisen, J., Klipstein-Grobusch, K., Benetou, V., Zavitsanos, X., Tumino, R., Galasso, R., Bueno-de-Mesquita, H. B.,

- Ocké, M. C., Charrondière, U. R., and Slimani, N., 2002. Variability of fish consumption within the 10 European countries participating in the European Investigation into Cancer and Nutrition (EPIC) study. *Public Health Nutrition*, 5 (6B), 1273–1285.
- WHO. 2011. Summary estimates of mortality for WHO member states for the year 2008 (Version April 2011) [Microsoft® Office Excel spreadsheet]. Available from: <http://apps.who.int/ghodata> [Accessed: 2012].
- WHO, and MSIF, 2008. *Atlas multiple sclerosis resources in the world 2008*. World Health Organization. Available from: http://www.who.int/mental_health/neurology/Atlas_MS_WEB.pdf [Accessed: 2012].
- Willer, C. J., Dyment, D. A., Risch, N. J., Sadovnick, A. D., Ebers, G. C., and The Canadian Collaborative Study Group. 2003. Twin concordance and sibling recurrence rates in multiple sclerosis. *Proceedings of the National Academy of Sciences*, 100 (22), 12877–12882.
- Willer, C. J., Dyment, D. A., Sadovnick, A. D., Rothwell, P. M., Murray, T. J., and Ebers, G. C., 2004. Timing of birth and risk of multiple sclerosis: Population based study. *BMJ*, doi:10.1136/bmj.38301.686030.63.
- Williamson, D. M., Henry, J. P., Schiffer, R., and Wagner, L., 2007. Prevalence of multiple sclerosis in 19 Texas counties, 1998–2000. *Journal of Environmental Health*, 69 (10), 41–45.
- Wingerchuk, D. M., 2012. Smoking: Effects on multiple sclerosis susceptibility and disease progression. *Therapeutic Advances in Neurological Disorders*, 5 (1), 13–22.
- Withers, C., 2009. xlutils: Utilities for working with Excel files that require both xlrd and xlwt (Version 1.4.1). Available from: <http://www.simplistix.co.uk/software/python/xlutils> [Accessed 2012].
- Wu, C., and Murray, A. T., 2005. A cokriging method for estimating population density in urban areas. *Computers, Environment and Urban Systems*, 29 (5), 558–579.
- Wu, G. F., and Alvarez, E., 2011. The immunopathophysiology of multiple sclerosis. *Neurologic Clinics*, 29 (2), 257–278.

- Wucherpennig, K. W., and Sethi, D., 2011. T cell receptor recognition of self and foreign antigens in the induction of autoimmunity. *Seminars in Immunology*, 23 (2), 84–91.
- Yang, Y., Lovett-Racke, A. E., and Racke, M. K., 2011. Role of IL-12/IL-23 in the pathogenesis of multiple sclerosis. *In: Neuroinflammation*. London: Elsevier, 107–136.
- Yenamandra, S. P., Hellman, U., Kempkes, B., Darekar, S. D., Petermann, S., Sculley, T., Klein, G., and Kashuba, E., 2010. Epstein–Barr virus encoded EBNA-3 binds to vitamin D receptor and blocks activation of its target genes. *Cellular and Molecular Life Sciences*, 67 (24), 4249–4256.
- Youinou, P., Pers, J.-O., Gershwin, M. E., and Shoenfeld, Y., 2009. Geo-epidemiology and autoimmunity. *Journal of Autoimmunity*, 34 (3), J163–J167.
- Yuan, Y., Smith, R. M., and Limp, W. F., 1997. Remodeling census population with spatial information from LandSat™ imagery. *Computers, Environment and Urban Systems*, 21 (3/4), 245–258.
- Zhang, R., and Naughton, D., 2010. Vitamin D in health and disease: Current perspectives. *Nutrition Journal*, 9 (1), 65.
- Zhernakova, A., van Diemen, C. C., and Wijmenga, C., 2009. Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nature Reviews Genetics*, 10 (1), 43–55.
- Zipf, G. K., 1946. The p_1p_2/D hypothesis: On the inter-city movement of persons. *American Sociology Reviews*, 11, 677–686.
- Zipf, G. K., 1949. *Human behaviour and the principal of least effort*. Cambridge, MA: Addison Wesley.
- Zivadinov, R., Iona, L., Monti-Bragadin, L., Bosco, A., Jurjevic, A., Taus, C., Cazzato, G., and Zorzon, M., 2003. The use of standardized incidence and prevalence rates in epidemiological studies on multiple sclerosis: A meta-analysis study. *Neuroepidemiology*, 22 (1), 65–74.
- Zuvich, R. L., McCauley, J. L., Pericak-Vance, M. A., and Haines, J. L., 2009. Genetics and pathogenesis of multiple sclerosis. *Seminars in Immunology*, 21 (6), 328–333.

[This page intentionally left blank.]