# Evaluation of vaccine effectiveness in older adults using routinely collected data: a quasi-experimental approach

Submitted by Adam Justin Streeter to the University of Exeter

as a thesis for the degree of

Doctor of Philosophy in Medical Studies

In September 2018

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

I certify that all material in this thesis 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:

# Abstract

Vaccination of older adults is a key component of public health policy, but further evidence is required to understand its effectiveness in practice. Electronic health records (EHRs) present a potential alternative to the gold-standard evidence of clinical trials, particularly for populations, such as older adults, who may be under-represented in trials due to ethical and practical constraints in recruitment. Importantly, EHRs also allow the real-world study of an intervention as it is delivered in practice, and its effect in clinically important sub-groups. However, EHRs are not purposed to collect informaton on confounders, which may bias results from the analyisis of routinely-collected data. This motivated my review of quasi-experimental (QE) methods as a means of indirectly adjusting for confounding. My published methodological review found that the longitudinal information available in EHRs offer many opportunities for mitigating against confounding bias, but many methods may be under-utilised. The prior event rate ratio (PERR) and its alternative formulation, described under the Pairwise framework, is a recently developed method that utilises longitudinal information. This before-and-after approach can be applied to rate and survival data, allowing an easy comparison to many trial results. The data on vaccination in UK older adults was also the basis for further study of the performance and limitations of the method beyond exisiting simulation studies. Through comparison to weighted regression, I demonstrated how the source of confounding and robustness of the results could be explored. In a novel application of the PERR and Pairwise methods to interactions, I investigated the effectiveness of the pneumococcal vaccine in older patients, and found evidence for an increase in effectiveness with age across the years of policy implementation, 2003-2005. In my investigation of the influenza vaccine in annual cohorts from 1997 to 2011, I found consistent evidence of a moderately protective effect against myocardial infarction, but that this may decrease with age. The evidence also indicated a protective effect against influenza itself, but no age trend in its effectiveness was detected.

# Table of contents

## Acknowledgements

I must thank my supervisors Professor William E. Henley and and Professor David Melzer for all their time, fantastic support and their expertise in helping me progress so well. I extend my thanks of course to my co-authors mentioned in chapers 2, 5 & 6 for their time and expertise, as well as Wendy Cowell, Sarah Vinnels and Leala Watson, who in their roles, also supported this work. Further thanks is due to Professor Stuart Logan, and PenCLAHRC, and also to Professor Willie Hamilton and Dr. Obioha Ukoumunne for their welcome feedback at the transfer viva.

On a personal note, I must acknowledge the huge support from my family: Anastasia, Oliver, Romy, and also from my late mother and her partner Doug.

# List of tables

# List of figures

## Author’s declaration

I declare this thesis and the work therein are my own original work and have been generated by me. The nature of the work meant that that some collaboration was required, and the co-authors are listed in chapters two, five and six. Clinical advice was received on the choice and identification of the outcomes as well as a review of the manuscript from the co-authors of chapters five and six. As systematic review, chapter two was necessarily a collaborative effort, with advice on, and input into, the literature search and a review of the text by the listed co-authors. However, I was responsible for the design and analysis of the studies, as well as interpretation of the results and the writing of those chapters and the thesis as a whole. I am grateful to the advice and assistance that I received from my co-authors and supervisors in helping me deliver this PhD.

# Abbreviations

| | |
|---|---|
| BSA | Bayesian sensitivity analysis |
| CAP | Community-acquired pneumonia |
| CPRD | Clinical Practice Reseach Datalink |
| DAG | directed acyclic graph |
| DCF | differential case fatality |
| DiD | difference in differences |
| EHR | electronic health record |
| FE | fixed effects |
| FV | ‘flu vaccine group |
| HES | hospital episode statistics |
| HR | hazard ratio |
| IPTW | inverse probability treatment weighting/weighted |
| IVA | instrumental variable analysis |
| IV-GMM | instrumental variable generalised method of moments |
| MCSA | Monte Carlo sensitivity analysis |
| MI | myocardial infarction |
| NCO | negative control outcome |
| PA | perturbation analysis |
| PERR | prior event rate ratio |
| PERR-ALT | alternative prior event rate ratio (also covered by maximum likelihood expression in the Pairwise framework) |
| PH | proportional hazards |
| PPV | polysaccharide pneumococcal vaccine |
| PS | propensity score |
| PSC | propensity score calibration |
| PT | perturbation testing |
| QE | quasi-experimental |
| QOF | Quality Outcomes Framework |
| RCD | routinely collected data |
| RCT | randomised controlled trial |
| RD | regression discontinuity |
| SA | sensitivity analysis |
| TVC | time varying covariates |

# Chapter 1 - Introduction

## 1.1 Introduction

It is the digital storage of routinely-collected health data that underpins this PhD as well as many other current, observational investigations into human health. Many casual observers will be aware of the rise of the “big data” phenomenon - an inevitable consequence of improvements in not only computer hardware, processor speeds and data storage, but also in the connectivity of distributed computing. This has greatly influenced the direction of health research, with benefits to the relatively new science of bioinformatics through increases in the capacity to store and analyse ever larger data from genetic studies and molecular biology. Simultaneously, the growth of “big data” has led to a greater investment in other observational studies, and the development of “real-world evidence”: Large, observational data on a population from routinely-collected electronic health records (EHR) may allow investigators to observe the effect of an intervention, away from the constraints of a clinical trial and according to how it might be delivered in practice, under real-world conditions. Extra information is likely to be available from longer follow-up times, enabling the monitoring of long term adverse events, or the discovery of unexpected, secondary benefits of a treatment. Furthermore, larger data captures more information across a wider mix of patients, and hence results can be compared across a variety of key sub-groups. EHR data tend to be more representative of the population from which they have been sampled, than trial data, and so results from analysis are more easily generalisable to that same population. In particular, some risk groups, such as older patients, may be under-represented in trials due to ethical and practical issues in recruiting to clinical randomised trials. However, in the absence of experimental rigor designed into clinical trials, this improvement in the representativeness of the sample (and thus, the generalisability of the results to the population of interest) may come at the expense of the accuracy of the results i.e: internal validity. The doubt over internal validity implies that results may not be reliable estimates of the true effect being investigated.

An oft cited example in relation to the problem of misusing “big data” for health is that of Google Flu Trends. Briefly, this approach used an unsupervised (i.e: theory free), algorithmic machine-learning to estimate the arrival of the annual influenza season based on the top 50 million search terms. The findings were published in Nature and the system proved to be much quicker than the usual surveillance through the sentinel laboratories of the Centre for Disease Control and Prevention. Unfortunately, the system also proved to be error-prone when the spread of influenza-like illness was overestimated by nearly a factor of two for the 2012/2013 influenza season. This was a problem of prediction modelling however, rather than one of inference, but the anecdote serves to illustrate the need for a robust model-based approach to investigating clinical questions.

The threat to internal validity may arise through more than one mechanism, but all variously lead to biased estimates of the treatment or exposure under investigation. The focus of this PhD was to seek causal inference from observational data, mitigating for the bias arising from an absence of information about important prognostic factors related to the outcome of interest. Given the prognostic factors are often likely to be imbalanced between exposure groups, confounding the effect of treatment, the resulting bias is referred to as confounding bias. Confounding and how it biases causal inference will be given a fuller discourse in a subsequent section in this chapter.

Methods are available for correcting for confounding bias, and many of these were presented in a comprehensive systematic review, which was performed as part of this PhD project and published in the Journal of Clinical Epidemiology. This is included as the second chapter of this thesis, and is supplemented with a subsequent chapter comprising a review of sensitivity-analysis methods, presenting a complementary approach to the problem of confounding.

As will become apparent in the method review, many methods were developed for economic data, and many are only applicable to continuous outcomes. Health research is often concerned with the epidemiological study of outcomes presented as binary clinical events, such disease states or death. The complement to the study of the rate at which clinical events occur is a prognostic investigation of the time until such an event occurs. In health research, survival until an event such as a disease or death or recovery from

disease is of intrinsic, prognostic clinical interest, and so survival analysis is often the method used for evaluating the effect of intervention in many trials. Therefore, this PhD sought to focus on extending the application of particular quasi-experimental (QE) methods that can be used to adjust for bias arising from unmeasured confounding in survival data. These methods were the prior event rate ratio (PERR) method, and its alternative formulation, described as under the Pairwise framework. Although these particular methods were covered in the method review of chapter two, a literature review of the PERR and Pairwise methods is presented and discussed in chapter four.

Following chapter four are two chapters, which use QE methods applied to EHR data, to investigate the real-world effectiveness of vaccination in older adults. Observational studies such as these have an important role in the study of older adults, and other populations that may be under-represented in clinical trials due to ethical and practical barriers to recruitment. The data used for the studies are introduced in a section below (section 1.6), and in another (1.7), details of the clinical question and the population are presented. Both chapters five and six have been written with intention of soon being submitted for publication in a clinical journal and will likely be submitted between the submission of this thesis and the viva voce. In the final chapter, the findings are summarised and the implications presented. The challenges in applying the PERR and Pairwise methods to this particular population and their performance are also discussed, along with any further methodological constraints and opportunities for development that were identified.

## 1.2 History of causal Inference

A definition of causal inference, seemingly so fundamental to empirical scientific discovery, has its roots in philosophy, and a history of development that dates back to work by John Stuart Mill (A System of Logic, Ratiocinative and Inductive, 1843), David Hume (A Treatise of Human Nature, 1739) and Sir Francis Bacon (Novum Organum, 1620), with influences from the classical Greek world in figures like Aristotle [1,2].

The concept of causal inference may at first seem intuitive to lay-people and scientists alike, as it asks the question:

*What causes an observed event?*

It is this experimental approach to resolving the problem of causal inference that is found in Jerzy Neyman's work on experimental design in agronomy for his Master's thesis [3]. This was considered by many to be a landmark development in a statistical framework for causal inference, although many leading statistical figures such as Ronald Fisher, Sir Austin Bradford Hill, William Cochran, David Cox and others have been acknowledged as contributing to the development of what came to be known as the potential outcomes framework [4–6]. The understanding of causal inference in the context of this PhD project was based on the potential outcomes framework, also referred to as the counterfactual model. However, it is accepted that this may oversimplify and reduce real-world mechanisms to an untestable theory [7].

The potential outcomes model is sometimes attributed to Donald Rubin [8] as Rubin's causal model [9], although Rubin, himself, modestly acknowledged that his efforts extended much work that had preceded his. While the model provided a framework for understanding the mechanism of inference from experimental data, this was also extended to observational data [10]. Donald Rubin with Paul Rosenbaum would later seek a solution to the problem of bias in observational data arising through from the imbalance in prognostic variables between treatment groups – the confounders of the effect of treatment. They

proposed that the unbiased average treatment effect could be estimated through either matching on, or adjusting for, a propensity score – the predicted membership of one of two treatment groups from a logistic model of the measured confounding variables [11]. Robins, Hernán and Brumback also demonstrated the application of the propensity score to the construction of the marginal structured model [12]. This estimated the marginal effect, the inverse probability treatment weight estimator, from an analysis weighted by the inverse of the propensity score. Stabilisation of the weights could also be achieved by adjustment in the numerator of the weights for the expected probability of treatment group membership, thereby reducing the influence of extremely large weights from a few subjects. However interpretation of the marginal estimate requires caution in the presence of consequential heterogeneity [13], where the estimated effect on the treated group will be different from that of the controls.

## 1.3 Causal inference framework

Central to the framework is Rubin's condition of ignorability, using notation in Rosenbaum and Rubin's work on propensity scores [11]:

$$(Y_0, Y_1) \perp A$$

which gives the joint independence of potential outcomes $Y_0$ and $Y_1$, and treatment assignment, $A$. Hernán deploys a useful and more succinct notation [14] for expressing causal effect in the context of potential outcomes. For the outcome $Y$, and a binary exposure or treatment $A$, each individual provides a pair of counterfactual outcomes $Y_a$:

$$Y_a \perp\!\!\!\perp A$$

This implies the property of exchangeability to be true so where the potential outcome $a$ of outcome $Y$ is independent of the actual exposure or treatment $A$. This may at first seem counterintuitive, but more simply it states that membership of exposure group $A$ should have no effect on the potential outcomes, $Y_a$, of each individual. Put another way, this means that for any individual, if $A$=1 is assigned, potential outcome $Y_{a=0}$ will equal the observed outcome under $A$=0, and vice versa.

Knowledge about the potential outcome of each individual, i.e: knowledge about *a=0* when *a=1* is observed and vice versa, renders the cause identifiable. Simply, causation cannot be attributed if only *Y*=1 is observed under *A*, since the counterfactual outcome *Y* of *A* could either be 1 or 0 [15]. Similarly the same argument is true, if *Y*=0 is observed under *A*. Hence causation can be inferred when:

$$Y_{a=1} \neq Y_{a=0}$$

However, the impossibility of observing both potential outcomes of each individual under *A* should be clear, so that it is impossible to estimate the individual causal effect, δ:

$$Y_{a=1} - Y_{a=0} = \delta$$

but, if the property of ignorability can be assumed, then causation can be inferred from association:

$$Y_a = Y|A$$

Under this property, the group comprising individuals under one level of the exposure are exchangeable with those under the other level. This enables the effect of exposure, say *T*, to be estimated from individuals that are different in each level of exposure. Therefore while the outcome for the exposed individual, $Y_{a=1}$, from set of individuals from population *A* under exposure level *T=1* is observed, its counterfactual can be observed as the outcome, $Y_{b=0}$, under exposure level *T=0* comprising a different set of individuals from population *B*. This property of exchangeability assumes E[$Y_{a=0}$] = E[$Y_{b=0}$] (and conversely, E[$Y_{a=1}$] = E[$Y_{b=1}$]) in the population of individuals. While the true average causal effect for individuals in group *A* is expressed as risk difference as

$$ACE = E[Y_{a=1}|T = 1] - E[Y_{a=0}|T = 0],$$

the observed effect is the difference between the outcomes in group *A* under *T*=1 and outcomes in group *B* under *T*=0, and can be expressed as

$$\widehat{ACE} = E[Y_{a=1}|T = 1] - E[Y_{b=0}|T = 0]$$

$$= (E[Y_{a=1}|T = 1] - E[Y_{a=0}|T = 0]) + (E[Y_{a=0}|T = 0] - E[Y_{b=0}|T = 0])$$

This is expressed equivalently, and perhaps more succinctly, as the outcome parameters of populations *A* and *B*, using Greenland and Morgenstern's notation[2]:

$$\mu_{A1} - \mu_{B0} = (\mu_{A1} - \mu_{A0}) + (\mu_{A0} - \mu_{B0})$$

where the true effect is $\mu_{A1} - \mu_{A0}$ , and $\mu_{A0} - \mu_{B0}$ is the potential bias between populations *A* and *B* (where $\mu$ denotes the population mean for the counterfactual of a particular group indicated by the subscript). The true effect can also be expressed analogously as a ratio:

$$\frac{\mu_{A1}}{\mu_{B0}} = \frac{\mu_{A1}}{\mu_{A0}} \cdot \frac{\mu_{A0}}{\mu_{B0}}$$

There are of course other assumptions required in the identification of causal pathways. The exposure or intervention has to necessarily be well-defined, and not subject to multiple interpretations or versions [16]. There has to be independence between individuals with respect to the intervention, and also the outcome, as in robust experimental design [17], referred to in the observational sense as the stable-unit-treatment assumption [18]. Mediation has not been discussed and in the discourse so far, has been assumed to be absent, yet the causal mechanism may not be properly identified without information on any mediators present [19]. This consideration links causal inference of the average causal effect to structural equation modelling, in which the interactions and the heterogeneity of the average causal effect are also considered. Causal inference in the presence of interactions was a topic of interest for this PhD project, and results from this are presented in later chapters. It should be noted in the presence of interactions, the exposure effect, when expressed as a ratio, will be different for the marginal effect and the conditional effect within strata of the covariate[2]. Whether the marginal and within-stratum conditional risk or odds ratios are equivalent is an issue of collapsibility. This sometimes gives rise to marginal and conditional ratios that are in the opposite direction, known as Simpson's paradox [20], which is dependent on the measure of association. This is sometimes mistakenly identified as confounding, i.e: missing information on covariate imbalance.

## 1.4 Causal inference and unmeasured confounding

The development of causal inference for observational data with methods based on the property of conditional ignorability assumes the availability of information on confounders within the data, so that individuals in the exposed and unexposed groups of *A* are conditionally exchangeable given confounder(s) *C*:

$$Y_a \perp\!\!\!\perp A|C$$

Without the tool of randomisation and experimental control over endogenous factors in a randomised controlled trial (RCT), unbiased causal inference from observational data relies not only on the correct specification of confounders in the inferential models, but also on the confounders , C, being identified and recorded in the data in the first place. Reliance on the availability of information on all confounders is likely to be unrealistic with retrospective observational data. Even when data collection can be planned for prospective studies, there may be practical difficulties in collecting data for all primary confounders, and a failure to either identify or correctly model the confounders leading to persistent bias [21,22].

There are different approaches offered by a raft of methods that could be best described collectively as quasi-experimental methods. Developed in parallel to those methods based on propensity scores found in medicine, QE methods largely have their origins in the disciplines of social sciences and econometrics [23–25], where randomised experiments are not practicable. A more comprehensive review of these methods is given in the published paper comprising chapter two.

The implication in the discourse on, and history of, causal inference by Cook and Campbell [26] is that the answer to the question regarding *what has caused an event* will depend on the philosophy that is followed in seeking the answer. In their own framework for causal inference, the distinction between “molar“ and “micromediation” is appropriate for medicine and biology, in which causality may be observed at the molar level (e.g: human physical activity leads to energy expenditure), but ultimately occurs at a micromediational level (e.g: the complex system of biochemical pathways involved between physical activity and energy

expenditure). Additionally, the comparison between open and closed systems helps underpin understanding of causal inference in medicine and human health. In observational data, the systems are decidedly “open”, easily leading to the “essentialist” view of cause and effect, in which all causal pathways are necessarily considered together. RCTs therefore attempt to emulate a closed system, in which stratification may control any endogenous relationships between known confounders and the outcome, and randomisation attempts to stochastically control for any remaining unmeasured confounding. The RCT therefore leads to a more experimentalist view of causal inference:

*What will happen if I change what I think is the cause?*

or perhaps more inferentially:

*Will changing what I think is the cause explain the cause of an observed event?*

While causal relationships need to be considered in designing an RCT to limit chance confounding relationships due to imperfect randomisation and/or non-compliance, the open system representing observational data relies more heavily on understanding the causal pathways for the correct estimation of cause-and-effect in subsequent modelling. In this context, Pearl [27] considered causal inference to be a “nonparametric generalisation of the linear structural equation models” first developed for research in Economics [28]. It was proposed that this could be helpfully illustrated through diagrammatic representation such as directed acyclic graphs (DAGs). Pearl considered an understanding of the underlying deterministic data-generation process to be essential for identifying true confounders as well as cause-and-effect [29]. Paradoxically, such an understanding may be difficult to achieve in the open system of observational data [30]. To that end, DAGs merely represent a hypothetical relationship, and the

risk oversimplifying causal pathways at the micromediational level of an open system.

After nearly a century since Jerzy's Master's thesis, it is surprising that the potential outcomes approach is still challenged today [31]. Questions have been raised over its relevance to description and surveillance within the spectrum of epidemiological investigation that extends to the aetiological [32]. While criticism of the formal approach to quantitative causal inference in epidemiology has been shown to be misguided [33,34], another debate has highlighted the danger of oversimplifying or mis-specifying causal relationships, which DAGs are merely purposed to illustrate [30,35]. Disagreements over the exact role of causal inference in epidemiology may in part be down to the RCT paradigm as the gold standard and exemplar of causal inference, which as already noted, may not sit well within the essentialist understanding of the open system that is human biology. Aalen, Roysland et al. perhaps best diagnose the role of RCTs in the confusion over causal inference by pointing out the limitations of RCTs in understanding causality: "Intervention and manipulation exhibit causality, but do not necessarily define it" [36]. This is understood to acknowledge the inadequacy of the one-variable-at-a-time approach under the idealised conditions of an RCT for understanding complex biological networks of cause-and-effect particularly under homeostatic equilibrium, where observed variables may act as mediators and moderators of other effects. Beside their excellent treatise on the development of causal inference, Aalen, Roysland et al. argued a compelling case for explicitly including direction of effect relative to time in the causal inference framework, a property that is missing from much of the literature discussing causation, including DAGs. They argued that the time direction of a relationship can simply determine whether or not to condition on a variable rather than relying on identification of variables as possible colliders. It is in this context of time in causal inference, that the argument is made for inclusion of longitudinal information in the QE adjustment of longitudinal data in the method review of chapter two, which in turn supports the review of the PERR and Pairwise methods in chapter four and understanding of the analyses conducted in chapters five and six.

## 1.5 Confounding bias and quasi-experimental methods

The problems posed by identifying potential confounders and correctly specifying the causal pathways of the confounders and other variables have contributed to the need for what is defined for the purposes of this PhD project as a QE approach to causal inference from observational data. Missing observations and missing information on key variables in the causal pathway are common problems with observational data. Variables that are associated with both the outcome and the exposure will confound any causal relationship between the outcome and exposure. (Paradoxically, if the confounder is measured without error and correctly specified in the inferential model then it is no longer a confounder, but just another adjustment variable. Therefore, the extra clarification of confounders as unmeasured or otherwise may be regarded as superfluous in the context of statistical modelling). In medicine, where interest is in the risk from exposure to a disease or in the effectiveness of a particular health intervention, the exposure or treatment is likely to be discrete and represented as a binary indicator variable. Therefore, potential confounders are any prognostic variable or a predictor of the outcome, which is distributed unequally between the exposed groups. Where the exposure or intervention is continuous, such as in the case of blood pressure or drug dosage, then prognostic variable would be distributed unevenly across the spectrum of the exposure to qualify as a confounder. Consequently, this creates uncertainty over how much of the difference in outcome between exposed groups should be attributed to the exposure itself or the confounding variable, which is wholly or partially aliased with the exposure. Age, which may commonly determine the prognosis of a disease in patients, may be balanced between intervention groups in an RCT either deterministically through stratification or stochastically through randomisation. Away from the controlled allocation of the intervention in an RCT, patients are likely to be selected for treatment on important prognostic variables such as age. Therefore in an observational study of any real-world scenario, the distribution of age may be different in each treatment group, leading age to potentially confound any observed effect of treatment. If the estimated effect of an exposure or treatment is not adjusted for confounding variables, then the effect of the confounding remains in the residual of any regression, biasing the estimates of the variables that are confounded. This

creates a threat to what is termed as the internal validity of the model and its estimates. If a variable is causally linked to only one of either the exposure or the outcome, then the variable is no longer an endogenous variable and a potential confounder, but an exogenous variable. If causally linked exclusively to the exposure, it can be considered an instrument of the exposure. If a causal determinant of the outcome only, then it is just another adjustment variable to be included in the regression model to reduce the error of the estimates.

QE methods obviate the need to identify and correctly model all confounding relationships, although such methods often require meeting certain sets of assumptions. The assumptions required of analytical methods, that can be used in comparative effectiveness studies, are detailed in the method review of chapter two. Additionally, each method requires the data to either be configured in a certain way, such as having a longitudinal dimension for before-and-after designs, or exhibit certain properties, such as the availability of instrument(s) of the exposure in instrumental variable analysis. Collectively when data conveniently provide such properties that may accommodate a QE approach, the data are said to provide a "natural experiment".

Applying QE methods to observational data to control for confounding bias obviously benefits the internal validity of the resulting estimates. However, there are also advantages to using observational data over experimental data, as will be explained in the following section on data and real-world evidence.

## 1.6 Routinely collected data

The aim of this PhD project was to research, apply and extend the use of particular QE methods to facilitate causal inference from observational data, primarily routinely-collected data (RCD) or administrative data. In medicine, this data is frequently encountered as patient information, collected and stored in the UK for the purpose of maintaining continuity of care. Since these data are routinely stored and accessed digitally, they are often referred to as "electronic health records" (EHRs), and henceforth will be referred to as such, although they may also be referred to as electronic medical records, electronic patient records and personal health records in the literature [37]. Such data may also include claims against health insurance, particularly in the US. Primarily

recorded for actuarial and administrative purposes in the reimbursement of claims by the insurance industry, these can also be considered as EHRs, since they track healthcare usage and have been used to facilitate studies into health and health service utilisation, as evidenced in the systematic review of chapter two.

The data for this PhD project were provided by the Clinical Practice Research Datalink [38], formerly the General Practice Research Database. The database collects primary-care data from GP surgeries using the Vision/EMIS IT systems, but began life as a useful record-keeping system in the management of a single General Practice [39]. With development by information technology specialists and linkage to data from other databases including the Office of National Statistics and Hospital Episode Statistics, this had grown into a data brokerage service, providing observational data for research into epidemiology, pharmaco-economics, pharmacovigilance and risk-monitoring [40] . A review of its resources estimated it had records on over 11.3 million patients from 674 practices [41]. Other sources of EHR data in the UK are Q-Research, ResearchOne, The Health Improvement Network and the Secure Anonymised Information Linkage databank.

More recently, the boundary between RCD and trial data has become increasingly blurred as large-scale trials are integrated with RCD [42–45]. The integration between trial and RCD is ideally suited for conducting pragmatic, open-label randomised trials. Dubbed point-of-care trials in the context of EHR data, the Randomised Evaluations of Accepted Choices in Treatment (REACT) trial recently explored the feasibility of using the CPRD system as a data collection service, and a real-time recruitment tool, randomising at the point of care [46]. This was in part motivated by the paradox of “research exceptionalism”, which describes the contradiction between guidelines informing clinical good practice and regulations governing trials [47]. Where there is an absence of sufficient evidence from comparative effectiveness studies, the choice in prescribing one out of set of similar drugs may entirely arbitrary and so the clinician may rely on his or her own judgement. However, should the clinician try to conduct research for determining the comparative effectiveness of the drugs, then paradoxically, this would be subject to rules that are far more stringent than those regulating clinical practice. The expansion beyond the merely

observational has been reflected in the changes to the services offered on the CPRD website, which now promises electronic Case Report Forms to facilitate point-of-care and phase III trials, in addition to the more commonly encountered use of EHR data for ecological, descriptive studies and for phase IV safety-monitoring. In parallel, there has been a commensurate growth in methodological innovation to facilitate conducting trials using EHRs, particularly in clustering to accommodate treatment allocation at the general practice level [48].

The shift in the use from descriptive epidemiological and risk-monitoring studies to including more inferential studies and trials has been supported in the UK by funding calls into research based on EHRs, primarily from a consortium of funding bodies led by the Medical Research Council, and also by collaborations with the pharmaceutical and healthcare industries, such as the EU-wide Get Real, a three-year project initiated in 2013. In the US, the greater integration of EHRs into research can partly be evidenced in legislative changes such as the introduction of the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 in the US. This mandated a process for improving the "privacy and security provisions" in the exchange and use of EHRs and set out a program for incentivising a "meaningful use of certified EHR technology". In the UK too, the use of EHRs are more and more an integral part of public health policy [49,50]. Overall, the view of stakeholders in the UK healthcare system would seem to vindicate the huge interest in using EHRs to improve clinical care, although there are some concerns about the quality of the data. From the report entitled *Future of Health: Findings from a Survey of Stakeholders on the Future of Health and Healthcare in England* [51], these concerns have tended to be over linkage to other useful data; uniformity and coherence between datasets; and reliability of the recorded data.

As witnessed by the view of stakeholders, it is perhaps a commonly held view on EHRs that there are many issues to resolve around linkage and data quality [52], and that next to RCTs, that the data are messy, plagued with missingness, noncompliance and the incorrect recording of observations. However, it was a view expressed by Tjeerd van Staa at the NIHR Statistics Group annual meeting in Sheffield 2016, that data are not necessarily error free because they are found on a trial's case report form (CRF). A CRF presents a pristine view of

the real-world mess that is evident through EHRs, and that the underlying data could so easily be messy and imperfectly collected. However, internal validation at the CPRD aims to ensure an acceptable standard of data quality [40,53]. The availability of up-to-standard (UTS) dates is one measure of assurance about the quality of the data. This is the date from which the data from a particular practice are determined by the CPRD to be of research quality based on the practice's death recording and gaps in the data. The UTS date was used in this PhD project in the selection of patient cohorts to exclude any patients registered to GPs, whose practices were not up-to-standard at the index date of the cohort. Likewise, patients were also excluded if the last-collection date for data at their GP's surgery fell short of the observed timespan for the cohort.

Although overcoming confounding bias is one of the main challenges of using EHR data, there are advantages to using these data for causal inference, and in trials in partnership with bodies like the CPRD as a broker in the recruitment of General Practices and provider of an electronic CRF service. The advantages of the large, observational studies based on EHRs are often compared to those based on RCT data:

1. EHRs typically provide much larger datasets than do RCTs, and the costs inextracting EHR data are typically lower too than the costs involved in running a trial.
2. The costs may also limit the length of follow-up for a trial, whereas for EHRs, once the data are abstracted, the follow-up lasts as long as the records for each patient.
3. Evidence from EHRs may complement existing trial evidence, especially where it may be unethical to randomise between treatment and control. Importantly, evidence from robust analysis of EHR data may also be used as pilot work for supporting future clinical trials.
4. A greater amount of information is also potentially available from EHRs across a wider variety of clinically interesting sub-groups than may be possible from a clinical trial giving rise to a more heterogeneous dataset, from which more patient-specific outcomes may studied
5. Besides larger sample sizes and affording greater statistical power, EHRs also potentially offers investigators the opportunity to observe the effect of an intervention as it happens in practice (so-called "real-world"

settings) away from the idealised settings of clinical trials. Clinical trials are often constrained by ethics and the practical considerations of recruiting to trial. Furthermore, patients may behave differently under observation in a trial than they would in real-world conditions (the so-called Hawthorne effect). Hence, the results from trials might not necessarily represent what might happen under real-world conditions. This may limit or even invalidate their generalisability, or external validity, or restrict their applicability to a much narrower section of the population than intended.

The last three points, and last two in particular, may be described as the basis for “real-world effectiveness”, which is a recurrent theme of this PhD project and other studies that use EHRs and routinely-collected data. It must also be considered that the very participation in a trial, patients may behave differently, in some reacting to the knowledge that they are being observed, giving rise to what is known as the Hawthorne effect. The absence of this effect is therefore implied in “real-world effectiveness”.

According to records, the controlled trial has a history extending all the way back to ancient history [54]. Notably James Lind is commonly acknowledged as being one of the first to conduct a controlled trial in recent history [55], but perhaps it is with the first randomised controlled trial of streptomycin in 1946 that truly marked the 20th century as the era of the RCT. RCTs, through design, randomisation and delivery, may reduce the impact of bias to deliver high quality evidence by anticipating and mitigating for information loss or imbalance of potential confounders. However, we should perhaps not underestimate the impact on medicine and clinical practice with the arrival of big data, and in particular RCD. Data rich sources such as CPRD may be utilised to broaden the scope beyond continuity of care for the individual to prevention and informing health policies at the population level [56,57]. It may be that the 21st century will be seen as the era of big data in medicine as well as in other disciplines. However, to maintain the quality of evidence, it is essential tools are developed to address the problem of confounding bias commonly associated with such data. It is towards this objective that this PhD project was directed.

## 1.7 Vaccination in older adults

As this PhD was in part supported by the NIHR School for Public Health Research Ageing Well programme, of particular interest was outcomes research in older adults, a population that has been identified as being under-represented in trials [58,59]. The reasons for their under-representation are multi-faceted, unclear and sometimes without justification, but at a practical level, impaired mobility and cognition may play a part in exclusion. The availability of EHR data on the population was therefore an ideal basis for addressing clinical questions that either cannot be realistically resolved through an RCT, or that seek real-world evidence of an intervention, whose efficacy may have previously been reported in an RCT. Studies, for which RCTs may not be appropriate, can be those investigating long-term adverse events and research into secondary unintended outcomes. Other possible investigations included research into the discontinuation of treatment and medication.

Vaccination against influenza and pneumococcal infection in older adults is a key public health policy intended to reduce the disease burden and associated public healthcare costs. An investigation of vaccine effectiveness would not only provide real-world evidence in this clinically important risk group, but would also supplement existing trial evidence for a clinically important risk group that may be under-represented in randomised trials due to ethical and practical barriers, such as gaining informed consent from older adults with dementia, or other cognitive impairments [58]. The scope for this investigation can be found in the protocol (Appendix C), which I co-authored as part of the PhD project, providing the major contribution to the statistical design and proposed analysis, and also the context in terms of current research. This protocol was submitted along wth that from another project using the same Gold Access to the CPRD data, and was approved by the Independent Scientific Advisory Committee of CPRD in December 2014. As the “oldest old” was one of the themes of Prof. David Melzer’s “Ageing well” investigations, the study on immunisation programmes in older adults was a logical extension of this package of work. This also presented an opportunity to build on the group’s work on polypharmacy in the elderly [60,61], and to complement the parallel work that had used the same data extraction [62].

A review of the literature on the vaccinations against influenza and pneumococcal pneumonia is incorporated in the chapters five and six in which the studies are presented. Further investigation of the real-world effectiveness of these vaccines was prompted by the problem of persistent bias that has affected previous observational research into vaccine effectiveness. Analyses relying on conventional regression models struggle to fully adjust for confounding bias, which is often unmeasured, and is compounded by decline in the functional status and age-related frailty in this older population [63]. The phenomenon of ageing and age-related frailty in older adults is well-documented [64–67], although presentation and causes are complex and multivariable [68,69]. Given the implicit vagueness of its definition, yet the complexity of its causes and effects, frailty could be thought of as a collection of complex latent variables. It is therefore uncertain whether enough of the variables, through which such latent effects may be manifest, could be identified and measured to control for confounding through an adjusted-regression approach to the analysis of nonrandomised data. As explained in previous sections, quasi-experimental methods may offer a way of adjusting, or rather mitigating for (given that no direct adjustment is made) unmeasured confounding.

A noticeable consequence of ageing and frailty is immunosenescence [70,71] - the age-related decline in immune function that is thought to explain the increasing susceptibility to infection from respiratory diseases, the vaccines for which were the subject of investigation in this project. It is immunosenescence and susceptibility to influenza and the pneumococcal infection that has directed the policy of vaccinating the older population. However, while vaccination may seek to counteract the age-related decline in the immune system, the problem of immunosenescence may weaken the intended immune response (immunogenicity) to the vaccine itself. The question of vaccine effectiveness in older adults is therefore comprised of two enquiries: is the vaccine effective in this population, and are there any trends with age that may suggest a weakened immune response to vaccination? This was the clinical question that motivated the study into the effectiveness of the pneumococcal vaccine in chapter five.

In chapter six, the effectiveness of the influenza vaccine was studied. However, rather than solely investigating its effect on influenza, greater interest lay in the potential benefits as a prophylaxis in preventing coronary disease. As will be discussed in the literature review of clinical findings, previous research has suggested that the influenza vaccination may be beneficial in helping to reduce rates of myocardial infarctions in the population of older adults. In seeking evidence for the existence of this effect, the study did not seek to explain the precise nature of the complex causal pathway that may exist between vaccination and outcome, but rather it acknowledged crucially the open system of causation that is human biology, as discussed in the previous section on causal inference.

## 1.8 Summary of objectives

The aim of this PhD project was to use routinely collected EHR data to evaluate the effectiveness of the pneumococcal and influenza vaccines in older adults. Retrospective recruitment to each study was different, reflecting the nature of the particular vaccine: The pneumococcal vaccine is considered to confer long-term immunity and the years, 2002 to 2005, over which the vaccine was introduced formed a natural experiment for this particular study. In contrast, older patients are recommend for revaccination against influenza every year, due to the annual changes in the virus mix and pathogen evolution, and so annual effectiveness was studied as far back as 1997, the year before the first wave in the introduction of the policy to vaccinate older adults.

Making use of the large data, the moderating effect of age on vaccine effectiveness was investigated in both vaccines. However, while the pneumococcal pneumonia was the primary outcome in the pnuemococcal-vaccine study, in the influenza-vaccine study, the primary outcome was myocardial infarctions, with influenza as a secondary outcome. In both interest also centred on detecting any change in effectiveness with age. However, as the data were observational, the studies also served as the basis for understanding the performance and limitations of a recently developed set of methods for dealing with confounding, called the prior event rate ratio (PERR) and Pairwise methods. This necessitated a full review of the methods, their

performance using simulated data, and their application in other studies. The method was also reviewed in the wider context of quasi-experimental methods. Succinctly, the aims of this PhD were summarised as:

- Conduct a systematic methodological review of QE methods and their application to longitudinal data as the context for the application of the PERR and Pairwise methods to the longitudinal data of EHRs, to understand the relative performance of each method and the strengths and weakness of using different QE methods that could be applied to EHR data [chapter 2]
- Conduct a review of sensitivity analyses, as a complementary approach to dealing with unmeasured confounding. The literature search for this was conducted in tandem with that for the methodological review [chapter 3].
- Perform a full review of the PERR and Pairwise methods, focussing on their relative performance from simulation studies, their assumptions and their application in studies since their development. Common settings using in subsequent chapters focussing on the methods' clinical application in vaccination studies will also be reported [chapter 4].
- Investigate the effectiveness of the pneumococcal vaccine in older adults using EHRs, and study effectiveness by age [chapter 5]
- Investigate the effectiveness of the influenza vaccine in older adults against myocardial infarctions and influenza. Again the effectiveness by age was also be investigated [chapter 6]
- In addition to the clinical findings from the vaccination studies, the relative performance of the PERR and Pairwise methods was also studied with a view to understanding the relative strengths and limitations that could lead to further methodological development [chapters 5, 6 & 7]

# Chapter 2 - Adjusting for unmeasured confounding in non-randomised longitudinal studies: a methodological review

DOI link: https://doi.org/10.1016/j.jclinepi.2017.04.022

Authors:

Adam J. Streeter[a,b]

Nan Xuan Lin[a,c]

Louise Crathorne[d]

Marcela Haasova[e]

Christopher Hyde[f]

David Melzer[g]

William E. Henley[a,*]

[a] Health Statistics Group, University of Exeter Medical School, University of Exeter, South Cloisters, St. Luke's campus, Exeter EX1 2LU, UK

[b] Biostatistics, bioinformatics & biomarkers group, Plymouth University Peninsula School of Medicine & Dentistry, University of Plymouth, Plymouth Science Park, Derriford, Plymouth PL6 8BX, UK

[c] Mathematics, physics & electrical engineering, Northumbria University, Sutherland Building, Newcastle-upon-Tyne NE1 8ST, UK

[d] Health Economics, University of Exeter Medical School, University of Exeter, South Cloisters, St. Luke's campus, Exeter EX1 2LU, UK

[e] Evidence Synthesis & Modelling for Health Improvement, University of Exeter Medical School, University of Exeter, South Cloisters, St. Luke's campus, Exeter EX1 2LU, UK

[f] Peninsula Technology Assessment Group, University of Exeter Medical School, University of Exeter, South Cloisters, St. Luke's campus, Exeter EX1 2LU, UK

[g] Epidemiology & Public Health, University of Exeter Medical School, RILD Building, RD&E Hospital Wonford, Barrack Road, Exeter EX2 5DW, UK

## 2.1 Abstract


### 2.1.1 Objective

Motivated by recent calls to use electronic health records for research, we reviewed the application and development of methods for addressing the bias from unmeasured confounding in longitudinal data.

### 2.1.2 Design

Methodological review of existing literature

### 2.1.3 Setting

We searched MEDLINE and EMBASE for articles addressing the threat to causal inference from unmeasured confounding in nonrandomised longitudinal health data through quasi-experimental analysis.

### 2.1.4 Results

Among the 121 studies included for review, 84 used instrumental variable analysis (IVA), of which 36 used lagged or historical instruments. Difference-in-differences (DiD) and fixed effects (FE) models were found in 29 studies. Five of these combined IVA with DiD or FE to try to mitigate for time-dependent confounding. Other less frequently used methods included prior event rate ratio adjustment, regression discontinuity nested within pre-post studies, propensity score calibration, perturbation analysis and negative control outcomes.

### 2.1.5 Conclusions

Well-established econometric methods such as DiD and IVA are commonly used to address unmeasured confounding in non-randomised, longitudinal studies, but researchers often fail to take full advantage of available longitudinal information. A range of promising new methods have been developed, but further studies are needed to understand their relative performance in different contexts before they can be recommended for widespread use.

**What is new?**

What is already known

- Unmeasured confounding is a threat to the validity of observational studies based on data from non-randomised longitudinal studies

Key findings

- Longitudinal information that can be used to mitigate for unmeasured confounding in observational data is not always fully or properly utilised in health research.
- Instrumental variable analysis and difference-in-differences were the most commonly encountered methods to adjust for unmeasured confounding in a review of the health literature.
- There are a range of promising new methods, some of which utilise longitudinal information to relax the assumption of time-invariance for unmeasured confounders, but these are yet to be widely adopted.

What is the implication?

- All available methods rely on strong assumptions and more research is needed to establish the relative performance of different methods for particular problems and empirical settings.

*Figure 1:* "What is new?" summary of contribution to research

## 2.2 Introduction

In the era of “big data” in medicine, the increasing availability of large, longitudinal patient databases is creating new opportunities for health researchers. A particular focus is on electronic health records (EHR) with routinely collected data collated from multiple care sites, often linked to external databases (e.g. death certificates). Built up over time, EHRs provide a sequential history of each patient’s encounter with the healthcare system. Examples of EHRs include The Clinical Practice Research Datalink (CPRD), The Health Improvement Network (THIN), QResearch and ResearchOne in the UK, and the Kaiser Permanente Northern California Oracle Research Database in the US. The value of large medical data recorded for administrative purposes in national registries is already recognised [72,73], with the provision of funds to expand the adoption of EHRs in research for patient benefit in the US with the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009, and in the UK, with a consortium of funding bodies led by the Medical Research Council. Another important source of information for health care analysis is databases of insurance claims, such as Medicare in the US, and in this review we do not differentiate between EHRs and claims data.

A strength of EHRs and claims data is that they make it possible to study the comparative effectiveness of interventions and the associated risk of side-effects in a real-world setting. Although randomised trials provide the gold standard of evidence, observational studies based on observational patient databases offer the potential to study more patients from a wider variety of risk groups with a longer follow-up period at a fraction of the cost. However, in the absence of randomisation, selection for treatment is often knowingly based on specific characteristics, such as frailty, disease severity or the risk of an outcome. If the indication for treatment is also related to prognosis, confounding by indication arises leading to biased estimation of effectiveness. There is a large pharmacoepidemiologic literature on this topic and current best practice is to use design-based approaches such as the Active Comparator, New User Design to help mitigate bias where possible [74].However, residual differences between the treatment arms other than the treatment itself may still confound the intervention effect under study whether or not such an approach is used. If

the confounding variables are both known to the study investigators and measurable, then these could potentially be adjusted for in prospective non-randomised studies. With retrospectively recruited subjects, however, the recording of such variables is outside the control of the investigator. Analyses of non-randomised studies that fail to account for relevant confounders may have important negative consequences for health policy and patient safety.

Methods described as the quasi-experimental (QE) approach [26], can be deployed to account for confounding by unobservable characteristics. These do not attempt to directly adjust for resulting bias, but use available information to achieve this indirectly under certain conditions and assumptions. The aim of this systematic review is to review current practices in dealing with unmeasured confounding in individual-level longitudinal health data and to capture methodological developments in this area. While previous systematic reviews have been conducted to look at use of propensity score methods for measured confounders [75,76], we are unaware of any systematic review comparing use of methods for addressing unmeasured confounding in non-randomised, longitudinal data. We were particularly interested in how an individual's history could be leveraged to evaluate the effects of unmeasured confounding and how the extra longitudinal information could be incorporated to improve adjustment for confounding bias. We intend for this review to contribute to the development of best practice in addressing unmeasured confounding in longitudinal data. The results should therefore help inform researchers intending to utilise "big data" from electronic health records.

## *2.3* Methods

### 2.3.1 Search strategy

Our search strategy was informed by, but not limited to, known methods for addressing unmeasured confounding. The search strategy is recorded in Appendix A – methodological review search terms. The following electronic databases were searched: MEDLINE (via OvidSp including In-Process & Other Non-Indexed Citations) and EMBASE (via OvidSp 1996 to 2015 Week 21). We included all citation dates from database inception to May 2015. All references were exported into Endnote X7 (Thomson Reuters).

### 2.3.2 Inclusion and exclusion criteria

The review included any non-randomised comparative studies that sought to adjust for unmeasured confounding in longitudinal data with repeated observations on identifiable individuals. In the interests of good practice, eligible papers had to explicitly identify the problem of bias arising from the selection on unobservable characteristics in the data, rather than routinely apply a QE design without this justification. For estimates of comparative effectiveness, eligible studies had to have independent control arms for each treatment of interest. Therefore, single arm studies were excluded. Studies based on case-only designs, including the case-crossover design and the self-controlled case-series design, in which confounding is controlled by making comparisons between exposed and unexposed periods for the same individual were also excluded. Observational studies were not excluded based on the exposure under study so studies into the effects of passive exposures (medical conditions, environmental exposures etc) were included alongside studies of both the intended and adverse effects of active interventions. We note that good proxies for unmeasured confounding, or observed variables that sufficiently describe a latent variable such as frailty, would be preferable to dealing with the bias resulting from unmeasured confounders. If suitable proxies are identified and recorded, then there are in effect no unobserved confounders and the proxies could simply be adjusted for in the analysis, obviating the need for methods to adjust for the unobserved confounders. For this reason, adjustments for proxies of unmeasured confounders, including high-dimensional propensity scores, did not fall within the scope of this study. To be consistent with the “big data” theme of EHRs, a minimum sample size of 1000 participants was applied. This also set a minimum condition for the application of Instrumental Variable (IV) and Regression Discontinuity (RD) designs stipulated in the Quality of Effectiveness Estimates from Non-randomised Studies (QuEENS) checklist. Finally, we only accepted analyses of individual level data. We were aware that some studies may use analytical methods, such as difference-in-differences that aggregate the data at a treatment-group level. We therefore only included those studies, in which the same patients could be tracked over the time-frame of the sample. Conversely, some methods, such as instrumental variable analysis, make no explicit demands for longitudinal data at

the patient level. However, we included such studies where the sample was based on the availability of patient-level longitudinal information, with a history possibly but not necessarily preceding the time of exposure. We did not discriminate between data sources, as patient-level data will often arise from medical insurance claims in the US, as opposed to clinically-purposed databases in other countries.

Only studies written in English were included. The following publication types were excluded from the review:

systematic reviews of primary studies.

randomised controlled trials

cross-sectional data

preclinical and biological studies

narrative reviews, editorials, opinions

### 2.3.3 Study selection

Studies retrieved from the searches were selected for inclusion through a two-stage process according to the inclusion/exclusion criteria specified above. First, abstracts and titles returned by the search strategy were screened for inclusion independently by two researchers. In case of doubt, the article in question was obtained and a subsequent judgement on relevance was based on the full article. Disagreements were resolved by discussion, with involvement of a third reviewer when necessary. Following the initial screening, full texts of identified studies were obtained and screened firstly by a single reviewer. In case of doubt, a second reviewer decided on the suitability of a paper. Where multiple publications of the same study were identified, data were extracted and reported as a single study.

### 2.3.4 Evidence synthesis

The details of each study's design and methodology and the key characteristics of the data source were tabulated and discussed. We present a summary of the methods we found that can mitigate for confounding, or its synonyms as unmeasured, unobserved, hidden or residual. We note the historical frequency

and context of the application of those methods, to comment on progress in causal inference and identify directions for future research.

## 2.4 Results

### 2.4.1 Included studies

Our searches returned 734 unique titles and abstracts, with 275 papers retrieved for detailed consideration Of the 275 studies eligible for a full-text review, 154 were excluded (see flow diagram: Figure 2).

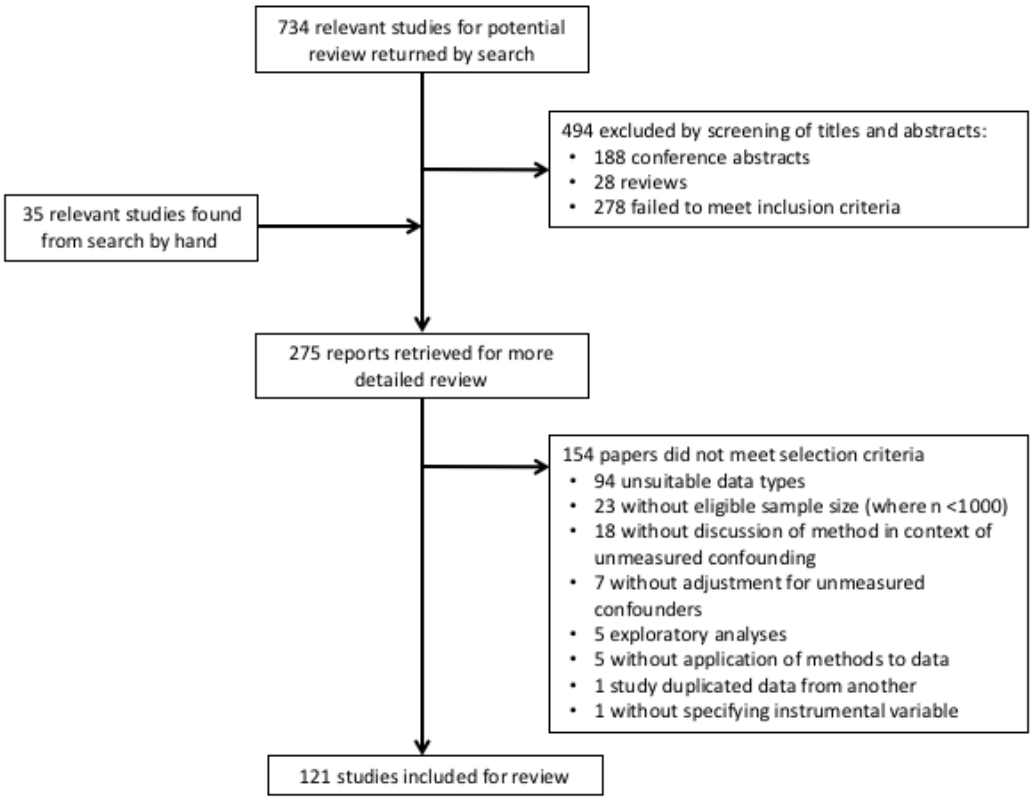


*Figure 2: Flow diagram for method review*

A total of 121 studies were identified as performing a QE analysis on non-randomised longitudinal data on human subjects, identifiable at an individual level, and so included for a full review of the text (Appendix B – table of studies included in the methodological review).

The QE methods identified in the review are summarised in Table 1. The most frequent method was instrumental variable analysis (IVA) found in 86 of the studies (Figure 3) – a method that uses an unconfounded proxy for the intervention or exposure. For successful adjustment, the proxy or instrument should be strongly, causally associated with the exposure or intervention, and the instrument should only affect the outcome through the exposure. In addition to IVA, three of these also applied difference-in-differences (DiD) – a method that typically uses pre-exposure outcomes to adjust for unmeasured confounding and assumes any trends unrelated to the exposure are the same in both groups. Seven more studies derived estimates from a combination of both IVA and DiD, two of which assumed an absence of higher order autocorrelation to use lagged observations of the treatment variable as an instrument. Beside the 11 studies applying DiD either in conjunction with or in addition to IVA, we identified a further 21 studies, in which the sole QE method was recognised as a DiD approach.

We found five studies applied the prior event rate ratio method, a before-and-after approach that can be aggregated to the treatment level for survival or rate outcomes and analogous to DiD. In all five cases the methods were applied to longitudinal, individual patient data. Similarly regression discontinuity (RD) was used for such data in three of the studies included for review. Another three focused on propensity score calibration (PSC). One study introduced perturbation testing and perturbation analysis, while another discussed the use of negative control outcomes.

| Method | Description | Obstacles to implementation | Frequency of methods |
|---|---|---|---|
| Instrumental variable analysis (IVA) | Upon identification of a suitably strong instrument, the influence of bias may be reduced through post-hoc randomisation. The instrumental variable should be highly determinant of the intervention or treatment received, while satisfying the exclusion assumption of being independent of the outcome other than through the treatment (Wright 1928; Angrist 1991). | In practice, finding an instrument with a sufficiently strong treatment association is a stumbling block in many analyses (Bound, Jaeger, and Baker 1995; Baser 2009). Association of the instrument with the outcome exclusively through the treatment is an untestable assumption, particularly if an indirect association exists through an unmeasured covariate. | 79 |
| Difference-in-differences (DiD) | A biased effect estimate between two treatment groups may be corrected by the same estimates from a treatment-free period prior to the exposure, which should be a measure of the confounding bias contributed to the treatment effect (Ashenfelter and Card 1984). Aggregated at the treatment group level, this is operationalised in regression as a period-treatment interaction. At an individual level, demeaning, first-differencing or dummy variables for each individual may yield bias-free fixed effects, contingent on assumptions. | The method is contingent on the availability of repeated outcomes in both periods and invokes a time-invariant confounding assumption: that the confounding bias as captured by the estimated treatment effect in a treatment-free period prior to exposure is constant through to the study period. | 24 |
| Prior event rate ratio (PERR) | Analogous to the DiD method for time-to-event or rate data, a biased estimate of the hazard ratio or the incidence rate ratio is adjusted through its ratio with that from a treatment-free prior period (Tannen et al. 2008). | As with the assumption for DiD, repeatable outcomes and a constancy of the unmeasured confounding bias is required across both periods, before and after the exposure. Prior event occurrence should not influence the likelihood of future treatment. | 5 |
| Fixed effects instrumental variable analysis (FE IVA) | IVA may be applied to DiD estimation to mitigate for second-order endogeneity: the time-varying part of the bias that may not have been adjusted for by DiD. | Assumptions of IVA apply | 5 |
| Dynamic panel model, or Instrumental variable - generalised method of moments (IV-GMM) | Lagged observations of the confounded (endogenous) explanatory variable are introduced in a first-differences fixed effects analysis so that the differences of the lags become the instrumental variables in a generalised method of moments estimation. | Assumptions of IVA apply. Here the differenced lags should not be correlated with the differences in the error terms. | 2 |
| Regression discontinuity (RD) | RD is a design for analysis based on a treatment assignment determined by a cut-off applied to a continuous variable that is preferably measured with some random noise (as many clinical tests may be). The outcome can then be modelled on treatment for individuals within a certain interval from the cut-off of the assignment variable to ensure exchangeability between individuals for robust causal inference (Thistlethwaite and Campbell 1960) | Where assignment is not sharply determined by the cut-off, an increase in the probability of treatment may be observed leading to a "fuzzy" version of RD. Continuity in the assignment variable is assumed, otherwise manipulation of assignment and reverse causality may be suspected. Assignment should be locally random around the cut-off and makes the weak assumption that no unobserved covariates are discontinuous around the assignment cut-off. | 3 |

| | | | |
|---|---|---|---|
| Propensity score calibration (PSC) | PSC adjusts for residual confounding in the error-prone main dataset by importing information about the unmeasured confounders from a smaller, external "gold-standard" dataset (Stürmer et al. 2005). Analysis in the main dataset is adjusted using a single dimension propensity score of the measured corrected for unmeasured confounding by regression calibration against the gold-standard propensity score. | Successful adjustment is wholly dependent on the availability of another dataset containing the exposure variable and error-free predictor, with individuals that are relevant enough to those in the main dataset and under similar enough conditions to assure sufficient overlap between the two datasets. | 3 |
| Perturbation testing/analysis (PT/PA) | This data mining approach aims to mitigate for unmeasured confounding by adjusting for many measured variables that are weakly associated with the unobserved confounding variables (Lee 2014). Simulation in the single reviewed example demonstrated this may require 100's, if not 1000's of perturbation variables (PV). | This requires a very highly dimensional dataset, which may ultimately obviate the need for indirect adjustment if the most or all of the confounders are captured. Simulation demonstrated the bias may be exaggerated if a confounder is inadvertently identified as a PV, requiring many more true PVs to correct the bias. The number of PVs may exceed the available degrees of freedom necessitating clustering. | 1 |
| Negative control outcome / exposure (NCO/NCE) | A negative controls causally related to measured and unmeasured confounders affecting the exposure and main outcome, but not directly causally related to exposure and outcome themselves. As such, the negative control may be used to detect confounding bias in the main study, and potentially to indirectly adjust for this (Richardson et al. 2014) | This assumes that the effect of the unmeasured confounders on the main outcome is similar to that affecting the negative control. | 1 |

*Table 1: Summary of methods to mitigate against unmeasured confounding captured by systematic review, and the frequency of their use amongst the captured papers*

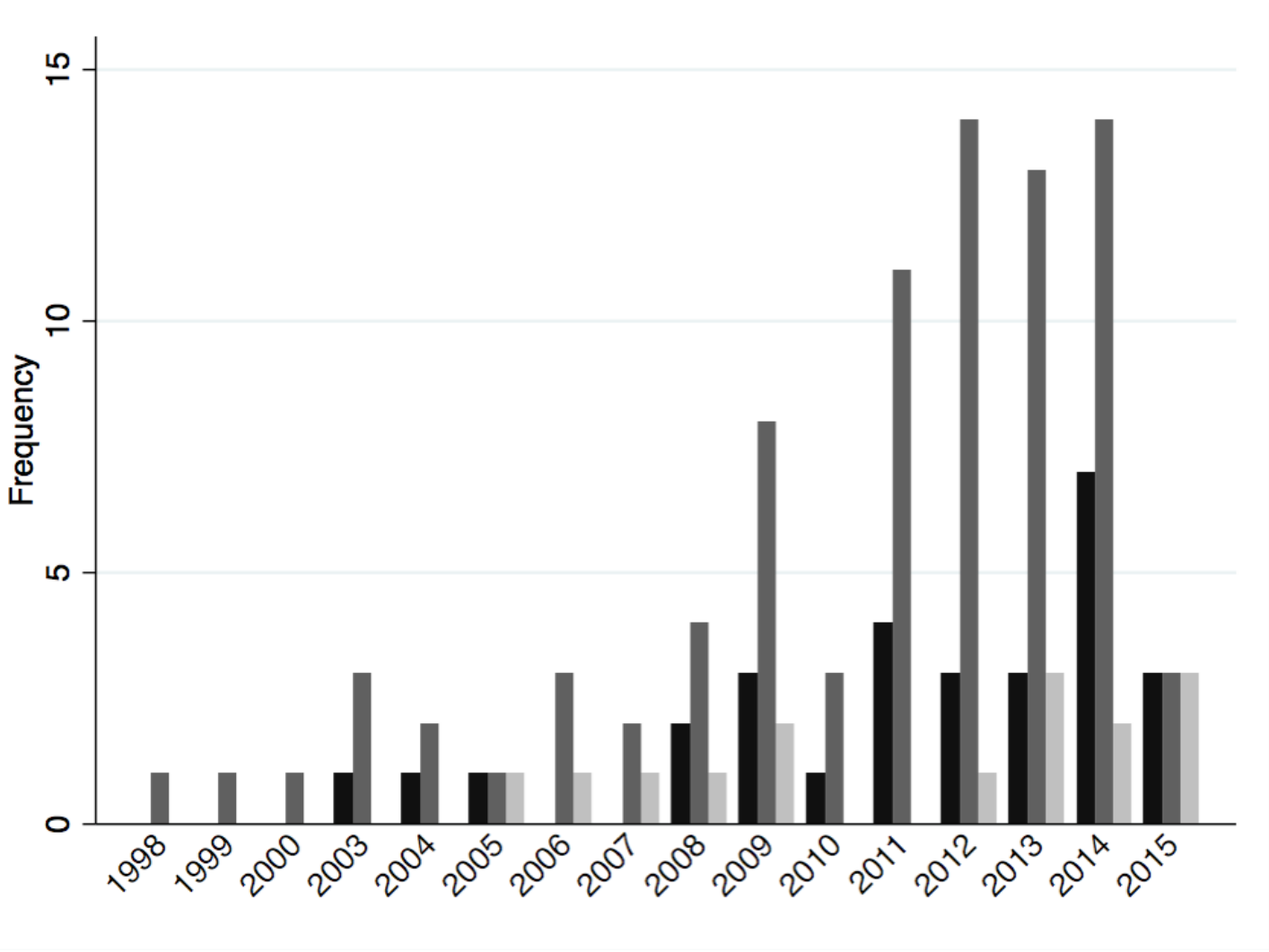


*Figure 3: Plot of frequency of reviewed methods for mitigating for unmeasured confounding by: difference-in-differences [black]; Instrumental variable analysis (IVA) [mid-grey]; Other [light grey] includes regression discontinuity, prior event rate ratio method, propensity score calibration, perturbation analysis, negative control outcomes, fixed effects with IVA and dynamic panel models. Note: the low frequencies in 2015 was attributable to the May cut-off for inclusion in that year*

### 2.4.2 Studies excluded at full text

The principal reason for exclusion in 94 of the studies, according to our eligibility criteria, was the absence of longitudinally observed, non-randomised outcomes on all individually identifiable persons, although other characteristics may also have justified their exclusion. No particular method was associated with the absence of longitudinal data on identifiable individuals with studies in this exclusion category comprising 59% DiD and 28% instrumental variable analyses compared, respectively, to 53% and 32% of all 154 of the rejected studies. Having fewer than 1000 longitudinally observed individuals excluded 23 studies, among which those using instrumental variable analysis (IVA) numbered 15. Seven were excluded for not employing a QE method for unmeasured confounding. Five studies presented exploratory analyses without a focused clinical question; five were either method reviews or commentaries without an application of methods to data; one study duplicated a dataset already marked for inclusion, while another failed to specify the instrumental variable used. Of particular note were the 18 studies using the DiD approach that were excluded because no explicit justification was made for using the method to address unmeasured confounding, or any of its synonyms. In these studies, justification of the method was centred more on econometric concerns over time trends, and presented in terms of controlling for those trends rather than pre-existing differences between the control and exposed group.

### 2.4.3 Results of the included studies

So far studies have been categorised according to their identified QE method. However, certain properties are shared across some of the methods, and can be classified according to how they reconcile their specific assumptions with the information offered by the structure of big, longitudinal data that typifies EHRs. In particular, we organised our results around how each method had incorporated longitudinal information, and the assumptions required. The stable of before-and-after methods, that includes PERR and DiD, implicitly incorporates longitudinal information. Thereafter the challenge is how to relax the assumption of time-invariant confounding. Conversely, IVA is not uniquely applicable to longitudinal data, but we were able to broadly classify the types of instruments used (Table 2), some of which

did utilise longitudinal information. We found out of the total 121 studies, 77 incorporated some element of longitudinal information into their analysis.

| IV type | Explanation/ Example | No. of papers | | | | Total frequency |
|---|---|---|---|---|---|---|
| Mendelian | Genetic characteristics :Single nucleotide polymorphisms | 11 | | | | 11 |
| Geographic | Differential distance between patient's postcode and nearest health facility | 19 | 1 | | | 21 |
| Time | Time-based characteristic of treatment such as date of therapy | 6 | | | 1 | 10 |
| Historical | Usually prescribing preference of physician or facility based on historical records of previously administered therapies | 31 | | 2 | | 34 |
| Lagged | Previous therapy or outcome of patient | 6 | | | | 6 |
| Randomisation | Original randomisation | 1 | | | | 1 |
| Other | Characteristics of individual e.g: age of patient, weight of offspring e.g: age of patient, weight of offspring | 8 | | | | 8 |

*Table 2: Frequency of instruments categorised by type used in instrumental variable analyses*

### *2.4.3.1 Incorporation of external/additional data*

The propensity scores (PS), the predicted probability of exposure or treatment conditioned on measured confounders,were used in the seminal work on propensity score calibration (PSC) by Stürmer to calibrate an error-prone PS against a gold-standard PS and hence arrive at an inference for the level of unmeasured confounding bias [77]. The two subsequent PSC papers examined the tenability of the

method's assumptions, firstly using simulated data to evaluate the conditions necessary to violate the surrogacy assumption [78]. The second primarily used simulated data and applied the results to registry data to demonstrate a framework for determining size and direction of bias from one measured and one hidden confounder [79].

### *2.4.3.2 High-dimensional data*

Since PSC collapses multiple, potential confounding variables down to the single dimension of a propensity score, the three PSC papers can also be considered a means of dealing with high-dimensional data. In addition to these, our review also included a novel data-mining approach that proposed to exploit the many factors (perturbations) that may be weakly associated with the unmeasured confounders from a high dimension dataset [80], for which longitudinal data may mitigate for incorrect adjustment of a collider. Perturbation analysis was successfully demonstrated on simulated data, although accidental inclusion of a measured confounder required many more perturbations to correct the resulting bias. Both the perturbation method and PSC were also proposed as sensitivity analyses.

### *2.4.3.3 Quasi-experimental adjustment without longitudinal assumptions*

Those studies characterised as using a QE method without any longitudinal dimension were PSC and PT as described above. We also added to this category 11 examples of Mendelian IVA [81–91] plus 32 other IVAs without historic or lagged instruments [92–123]. While time-based instruments may at first seem longitudinal, these instruments, such as date of therapy, would need to be related to previous exposures or outcomes to be considered longitudinal. In some cases, survival times or rate data were used, but such outcomes do not intrinsically imply longitudinal adjustment for confounding. In spite of these "cross-sectional" approaches, all studies were based on some form of longitudinal data at the person level, as demanded by our inclusion criteria. Among the 43 non-Mendelian IVA papers in this non-longitudinal category, one study adjusted for non-longitudinal fixed effects within twins [101]. In another three, discussed below, the analysis was supplemented with DiD [100,110], and with IVA applied to first-differences [124].

One study examined the effect of lagged, cumulative exposure to radiation on lung cancer in uranium miners and nuclear workers [125]. The problem of unmeasured

confounding was addressed using a method developed in earlier work that proposed negative control outcomes and exposures as a means of both detecting and potentially resolving confounding bias [126]. Here the choice of death due to chronic obstructive pulmonary disorder as a negative control outcome was informed by clinical knowledge of there being no direct relationship with the exposure except through the possible confounder, smoking. Given a plausible negative control outcome or exposure, the method offers at least a means of testing for confounding, and potentially a method of adjustment under the assumption that the association between the unmeasured confounder and the negative outcome is similar in magnitude to that between the same confounder and the outcome of interest.

#### *2.4.3.4 Quasi-experimental adjustment assuming time-invariant longitudinal information*

We found 36 IVA studies that used lagged information or history about the individuals' exposure as instruments [124,127–162]. One study had recourse to the random assignment from a previous study, and used this as an instrument [131]. Except for that and four other different exceptions, the instruments were all based at least in part on the previous intervention, or history of interventions, of the clinician or healthcare facility. Characteristics of the clinician or facility may be chosen as instruments as they are more likely to affect the treatment only. This avoids direct associations with the individual and their outcome, and so better enforces the exclusion restriction – the exclusion of the instrument's association with the outcome except through the treatment under study. While no assumptions are made about the dependence of confounding on time, the strength of the instrument clearly rests on a significant association between previous treatment(s) and the current treatment under investigation. In this regard, if the strength of an instrument varies with time, this may undermine its utility.

In total, 24 studies also incorporated longitudinal information through the stable of methods that, in an abuse of terminology, we collectively referred to as the DiD approach. These included the 18 examples cited as using DiD regression [163–180] alone, and four fixed effects (FE) [181–184]. Either through fixed effects at the individual level or through aggregate-level regression operationalizing the DiD approach, these methods "ignore" the effect of confounding, which is assumed to be time-invariant. At

the individual level, time invariant confounding can be ignored by assigning nuisance dummy variables for each individual, or cancelled out through demeaning the observations, or through the first differences of observations on each individual. Two of the studies also extended DiD to allow different exposure effects and trends across two-level sub-groups in the higher-order contrast of difference-in-difference-in-differences [168,173]. Fourteen studies also adjusted for individual-level fixed effects either through direct inclusion of their covariates, or through matching or weighting on the propensity score of the covariates. This was perhaps a more rigorous and precise approach, accounting for known confounders, and yielding smaller standard errors for the estimated treatment effect. However, an assumption of time-invariant confounding was still required, with a null difference between exposure groups in the prior period being evidence of adjustment for time-invariant confounding only. Two of the 24 DiD studies also re-analysed their data using IVA [100,110], which provided an albeit limited opportunity to compare the relative performance of these methods. In the study by Schmittdiel et al.  of how statins delivered by mail order affects cholesterol control [110], the intervention coefficient from modelling the single main outcome was larger through DiD analysis and its standard error smaller than those from IVA, large standard errors being a feature of weak instruments. The study by Lei and Lin investigated the effect of exposure to a new medical scheme on 15 health outcomes and rates of health-service utilisation [100]. The effects were either not significantly different from the null or were significant and of similar magnitude with similar standard error except for two outcomes, where the effect size was significantly larger for IVA.

Time-invariant confounding, also known as the parallel trends assumption, was relaxed by including dummy variables for the year and its interaction with the treatment dummy in a fixed-effects analysis, which allowed the unobserved trend to vary between exposure groups [183] using methods developed in economics and therefore not captured by this review [185,186]. The results from this DiD with differential trend model were presented alongside those from the simple pooled DiD model and DiD with individual fixed-effects for the effect of financial incentives in care services. Tests confirmed parallel trends could be assumed in three outcomes, but out of the five outcomes presented, four were statistically significant and in all, the estimated effect size by differential trends was greater.

Our review also included six studies applying the prior event rate ratio method, a before-and-after analogue applicable to survival and rate data [187–192]. The first two published were the seminal presentation of the method applied to registry data. Also included was a comprehensive evaluation by Uddin et al. of the performance of PERR under a wide array of simulated, theoretical settings, under which bias was shown to increase with a greater effect of the prior events on subsequent exposure or intervention. When prior events strongly influence the likelihood of treatment, the exposure effect from the PERR method can be more biased than estimates from conventional methods[121]. The problem was re-examined in a recently published study, which provided a more general statistical framework for PERR adjustment and considered the potential for generalising the method to allow more flexible modelling [192].

#### *2.4.3.5 Dynamic, longitudinal quasi-experimental methods and time-varying information*

While regression discontinuity (RD) could suggest a longitudinal design, this is not exclusively so, and two RD studies were excluded because of this (one applied to spatial data while the other data was not longitudinal). Of those included all three could be said to accommodate time varying trends [193–195], and two of these were nested within a pre-post design: Zuckerman et al. were explicit in their methodological study in identifying the robustness to time-varying confounding, in which inhaler use in asthmatic patients was served as both the outcome variable in the post-test period as well as the assignment variable in the pre-test period [195]. In the study of the effect school-leaving age on mortality by Albouy, different slopes were modelled for the assignment variable, year of birth, after the cut-off date [193]. This acknowledged different maturation rates after assignment. However, as long as the assumptions of the method were met, assignment should have been as good as randomised, and so no further assumptions about the temporality of confounding was required.

We also picked up six examples where IVA had been combined with either DiD or a fixed effects model, first appearing in our review with example from 2003 [196]. In Fortney's 2005 study of treatment for depression [197], this combination method was justified as a control for time varying confounding, referred to as second-order

endogeneity. Further examples of the fixed-effects instrumental variable model were found [198,199]. The roles of lagged treatments and outcomes as possible IVs and predictors were extensively considered in O'Malley's study of whether the introduction of more expensive medication could have led to improved cost-effectiveness in the long term [124]. The author cautioned that the exclusion restriction may be difficult to satisfy when using the lagged treatment as an IV after first differencing. However, two studies [200,201] used differences in the lagged explanatory variable as the IVs to adjust for second-order endogeneity in a first-differences analysis following methods, not captured by our review, but developed in the realm of Economics [202–204]. Referred to as the dynamic panel model or IV-GMM, this method was implemented efficiently through generalised method of moments. In their report on healthcare expenditure in patients with rheumatoid arthritis, Kawatkar et al. found the yielded estimates were further from the null with larger standard errors when compared to those from FE alone [200].

### 2.4.4 Implementation of methods

While choice of method in each study often rested on which extra information was available to address the issue of unmeasured confounding, method selection may also have been informed by the research area. The negative control method had its origins in epidemiology, with applications to occupational health policy. Likewise, the PERR method was developed exclusively on health data, with applications to drug safety and public health policy. Reflecting their origins in health econometrics, some studies were published in journals partially or entirely dedicated to the subject, with 15 published [100,124,163–166,173,176,181–184,196,197,200] in this field out of the 32 studies using DiD and 29 [93–95,98,103,104,109,111,112,114,116,120–122,128,131–134,140,144,147,150,205] out of the 86 using IVA. Under the inclusion criteria, all studies had health outcomes or interventions. Mendelian IVA necessarily includes genetic information, and all were published in health-related journals. In contrast, all three studies using RD were published in health econometric journals.

Before implementing one of the proposed methods, a natural first step is for the researcher to try to assess how much bias from unmeasured confounding is likely to be present. While many of the included studies reported raw or unadjusted

descriptive estimates, bias estimation was limited either to considering the contribution from known confounders, including those summarised as a propensity score, or to methods, such as perturbation testing/analysis and negative controls methods, in which bias evaluation is an incremental step in adjustment. Under the assumption of time-invariant confounding, the difference-in-differences method may potentially offer a way of evaluating bias by modelling group differences in the pre-exposure period. However, few studies evaluated hidden bias in this way [110,174,182]. The regression formulation of the DiD method effectively by-passes separate analysis of the prior period. Instead studies often discussed the within-group changes over time. Similarly, the prior-period estimate from the PERR method implicitly offers an evaluation of confounding bias under the same assumptions, yet none of the studies presented information on outcomes in the prior period in this way. A direct evaluation of unmeasured confounding is less straight-forward in IVA, with further diagnostic tests only recently developed for the association between instrument and confounders [206,207] .

## 2.5 Discussion

This review examined the application of methods to detect and adjust for unmeasured confounding in observational studies, and was motivated by recent calls to utilise EHRs. Most of the reviewed studies used more established methods such as DiD and particularly IVA. We summarised how studies exploit the longitudinal information afforded by EHRs.

It may be tempting to view electronic health records and medical insurance claims data as a problem of large observational data, and hence search for solutions through data mining. However, ethics governing patient data collection, plus limited clinician time is likely to preclude data with very large dimensions. For that reason, it is doubtful there would be enough dimensions for a method like Perturbation Analysis (PA) to be a practical solution. In addition, a greater number of variables would likely include enough information about the confounders to obviate the need for further adjustment through PA. More generally, the purpose of EHRs primarily as an administrative tool limits the scope for data mining of known confounders.

Similarly, limited availability of gold-standard datasets may have confined the use of external data, as in PSC, to but a few examples.

We were surprised by the number of studies using IVA alone. While Mendelian randomisation has its advantages for many studies as a reasonable guarantor of the exclusion restriction, in general IVA typically suffers from the weak-instrument problem, resulting in large standard errors and wide confidence intervals. Longitudinal data offer an opportunity to reinforce the exclusion criteria by choosing historical or lagged instruments. However in doing so, the causal structure needs to be understood to avoid opening up “back door” paths and inducing further bias [124]. DiD arguably offers advantages over IVA in being more intuitive and easier to conceptualise, and with the longitudinal data in EHRs it should be inherently easier to work with prior observations than to identify strong instruments. Even though before-and-after methods are not subject to the imprecision of weak instruments, the resulting estimates are only unbiased if the unobserved confounders exert a constant effect over the observation windows before and after exposure. Where multiple observations per individual exist, time may be paramaterised and different trends between exposure groups can be accommodated in DiD with differential trends, but a time invariant assumption about confounding must still be made. To partially or wholly relax this particular assumption, instrument variable analysis can be incorporated into the fixed effects model. Assuming the instrument’s exclusion restriction is satisfied then this doubly-robust approach affords the advantage of DiD over possibly weak instruments, while mitigating for some or all of the time-dependent confounders ignored by DiD alone. Similarly, where multiple previous treatments or exposures are recorded, the differenced lagged treatments can be utilised as IVs in a fixed effects model to accommodate time-dependent confounding bias using the generalized method of moments system, referred to as IV-GMM or the dynamic panel model.

Another potentially robust approach to unmeasured confounding would be the RD design, although the small number of examples in our review probably reflects the limited number of scenarios where this can be reasonably applied. Another concern over and above the usual technical challenges of applying the RD method is that in spite of heath records promising ample data, the sample would need to be reduced to an interval around the cut-off that ensures exchangeability of the two treatment

groups. In this case generalisability would be restricted to individuals with characteristics found in the interval. As with RD, PERR was another method that was found in relatively few studies. This may have been in large part due to its recent development, rather than any technically demanding aspect of its application, since it simply extends the before-and-after approach of DiD to survival and rate data - outcomes that are common enough in health research. However, the PERR approach does require strong assumptions including time-invariant confounding and the absence of an effect of prior events on likelihood of future treatment [192].

Methods such as IVA and DiD have their origins in the sphere of econometrics, where randomised experiments are rare. We found that in importing DiD, some of the studies failed to explicitly acknowledge the problem of confounding bias. Instead justification for the method was presented in terms of the common trends assumption. Discussion of possible confounding bias is regarded as essential by most QA toolkits for observational data, and it is important that health researchers explicitly recognise this threat to the internal validity of non-randomised studies. Conceptually a non-temporal analogue of DiD would be the NCO method, which itself was presented foremost as a method for detecting unmeasured confounding. Given doubts over satisfying necessary assumptions for their implementation, authors of this method along with propensity score calibration and perturbation analysis have suggested that, as sensitivity analyses, these can at least offer an insightful complement to QE adjustment.

Choosing between methods to reduce unmeasured confounding bias is challenging and we found few studies that directly compare methods. The performance of different methods will depend on factors such as the nature of the underlying confounding, the type of exposure and outcome, and the sample size [208] The type of data available will also guide the choice of method. For example, the instrumental variable method requires a suitable instrument and DiD / PERR require data on at least two periods. In practice, no one method is likely to be best suited to all problems, and it is essential for investigators to carefully assess the potential biases in each proposed study, where possible tailoring the methods or combination of methods to address these biases [209]. Our review has highlighted how use of longitudinal information is one additional and potentially important consideration in this process.

While our review focussed on the problem of adjustment using analytic methods, many problems associated with observational data may be pre-empted by use of an appropriate study design [210]. Before choosing an appropriate analytic method, it is recommended that investigators carefully identify and match individuals for the control and intervention groups in order not to exacerbate any bias [74]. The importance of study design is often discussed with a view to minimising confounding bias from unmeasured sources, with the subsequent adjustment accounting for observed confounders only [211], usually through the matching, weighting or adjustment of propensity scores [212]. Where the success of the design remains in doubt, or its criteria cannot be fully met, then investigators will inevitably need recourse to some of the alternative methods reviewed in this report.

The reviewed studies did not seek to distinguish between the different mechanisms of bias. Confounding by indication, deemed intractable by many researchers using the observed data [213], was seen to create additional sources of bias in two separate simulation studies applying the "longitudinal" method of PERR, when an association was modelled between prior events and treatment status in the study period [121,122]. Another common form of selection bias in pharmacoepidemiologic studies is the healthy user bias and this works in the opposite direction to confounding by indication, distorting treatment-outcome associations towards the treatment looking beneficial[3]. Further research is needed to understand how each of the methods in this review is affected by the different types of confounding.

An inherent limitation of this large, wide-ranging review is that it precluded meaningful data synthesis due to the mix of different data and study types. Furthermore, we could only find a few examples where the performance of different methods was compared within the same study. We also stipulated in the inclusion criteria that unmeasured confounding, or any of its synonyms, should be given as justification for methods in its adjustment. This may have inadvertently excluded some papers, where justification was implicit, but good practice in health research demands acknowledgement of this source of bias where applicable. While our search terms were specific to the scope of our review, we accept that this may have inadvertently excluded relevant methods and studies. Some methods, such as negative control outcomes, that were identified in the original search were not included as explicit terms in the search strategy, and further secondary searches

may have uncovered additional studies using these methods. We also acknowledge that there may be other relevant methods for addressing unmeasured confounding that have been missed by the search strategy. Consequently, we made inferences about the relative application of methods with caution. However, we were surprised so many studies focussed solely on IVA as the sole means of adjustment. A similar conclusion was echoed by a different review on regression discontinuity designs that found interest was growing in RD only as recently as 2014 [214].

By choosing to focus on methods with an independent control arm for each treatment, our review excluded case only designs including case-crossover designs (CCO) and the self-controlled case-series design. This class of methods addresses unmeasured confounding by making comparisons within individuals so that each individual acts as his or her own control. Another case-only design, the case-time control design, is an extension of the CCO design that uses information from a historical control group in a similar way to the PERR method. These approaches are reviewed by Uddin et al [208] and Nordmann et al [215].

This review has considered a range of promising new methods for addressing unmeasured confounding in non-randomised studies. However, consistent with prior research on dissemination and uptake of statistical innovations[146], the rate of knowledge translation has been slow and we found that most studies in our review used established methods such as IVA and DiD. A recent study by Cadarette et al has shown how Rogers' Diffusion of Innovations model can be used to describe the adoption of novel methodologies in pharmacoepidemiology[147] and this provides a useful resource for interpreting the uptake of methods in this review. Cadarette et al proposed five principles for authors of methodological innovations that may improve translation into practice [147]: (1) clearly describing the methods using foundational principles; (2) comparing results to established methods; (3) providing sample data, code or calculation examples; (4) early communication, support and testing; and (5) providing methodological and reporting guidance. These recommendations offer a useful checklist for researchers developing methods for addressing unmeasured confounding in observational studies. Of particular relevance in the context of this review is the need for more extensive evaluation and comparison of the emerging methods in a range of settings. The review also addresses the need for methodological guidance through highlighting the potentially important role of

longitudinal information in addressing confounding bias and has identified this as an area for further development.

## 2.6 Conclusions

Our review showed how seminal work in econometrics has influenced practice in dealing with unmeasured confounding in clinical and epidemiological research. Although the issue of unmeasured confounding is widely acknowledged, we found that longitudinal information in observational studies appears under-utilised. Lagged and historical characteristics associated with the treatment may help enforce the exclusion restrictions of instrumental variables under the appropriate causal structures, while before-and-after methods, such as DiD and PERR, afford an intuitive approach without the imprecision of weak instruments. Furthermore, they offer a direct evaluation of time-invariant confounding bias. The most robust methods we found applied instrumental variable analysis to the fixed effects difference-in-differences method, where such suitable instruments or difference lagged variables could be assumed to satisfy the exclusion restriction. While there are sometimes good technical reasons for choosing one mode of analysis over another, many questions remain over the most appropriate methods. All methods rely on assumptions, but little guidance is available to applied researchers as to the empirical settings in which particular methods can be safely used. Few studies directly compare different methods and more research is needed to the establish the relative performance of the methods in realistic settings.

# Chapter 3 - Sensitivity analysis for addressing unmeasured confounding: a methodological review

## 3.1 Introduction

Both SA and QE methods take an indirect approach to the problem of confounding. The QE methods of chapter 2 make an adjustment, or rather a mitigation for unmeasured confounders, by invoking assumptions and utilising other available information, such as longitudinal observations available on patients in EHRs. Where uncertainty exists over the tenability of the assumptions or the precision of the estimates, such methods can be presented as SAs to postulate how much of the observed effect could potentially be attributed to residual confounding. In this way, the distinction between QE methods and SAs is not entirely clear. Therefore, it was important to also consider QE methods in the context of SAs, and so the search terms for SA were also included in the search strategy for methodological review of chapter two (Appendix A).

Typically, an SA is either performed over a range of plausible settings for the confounding effect, or empirically discovers the degree of confounding required to move an observed effect to the null. Information on the direction and strength of association between the confounder and the outcome and key explanatory variables may be inferred from the data, or be imported from a source other than the dataset under analysis. If a high degree of imbalance of the confounder(s) across the treatment groups, or an implausibly high association of the confounder(s) with the outcome, is needed to change or explain treatment effect, then the results can be assumed to be reasonably robust to realistic levels of unmeasured confounding. Cornfield et al. [216] are often cited as the authors of seminal work on this approach. In this respect, adjustment methods and SA are very much complementary approaches to the problem of unmeasured confounding.

## 3.2 Methods

The search strategy for the SAs followed the same as that for the methodological review, already outlined in chapter two. The study selection process also followed that of the review. The search term for sensitivity analysis can be found amongst the terms for the methodological review in Appendix A – methodological review search terms.

### 3.2.1 Inclusion and exclusion criteria

The definition and implementation of SA can vary from study to study. For the review an SA was defined to be a method by which a conventionally adjusted estimate for an intervention is challenged by the introduction of a hypothetical confounding variable(s) for a range of associations between the outcome, exposure and confounder(s). This did not include covariate substitution or restriction to a range of covariates or sub-group analysis, approaches which were excluded from the review. An additional consideration in managing the size of the review results was the expectation that the results from the literature search would be dominated by observational studies routinely following good practice in applying SA to test the robustness of results from inferential models. For that reason, I focussed on studies that sought to either develop SA methods using either simulated or observed data of any size, or explicitly focussed on the application of SA.

## 3.3 Results

In all, 23 papers on sensitivity analysis were eligible for review, including three studying propensity score calibration (PSC), one on perturbation testing and one on negative control outcomes (Table 3). Some studies were based on seminal work not captured in the literature search for this review, which were nonetheless included for reference and discussion. One such was Greenland's review of SA [217], which applied the basic formulation for the "external† adjustment" of a single binary confounder affecting the odds ratio of the association between a binary exposure and outcome. This involves choosing a plausible range over which to vary the confounder-outcome association and the odds of confounder prevalence in each exposure arm, but

† "external" denotes manipulation by the author rather than external data

requires simultaneous interpretation of the confounder parameters. Margolis et al [218] deployed this method and compared it to another by Rosenbaum [219], which summarised the imbalance of confounder prevalence as single parameter to predict the observed outcome due to the exposure. Margolis et al. advocated a combination of both approaches to reduce the number of computations. A study by Cabral captured in this review [220] also followed this principle.

In their study, Arah et al [221] provided simplified formulae for uncontrolled confounding bias for different effect scales (risk difference, risk ratio and odds ratio) under a common framework, which can also accommodate polytomous unmeasured confounders. Formulae for SAs were subsequently extended to interaction analysis by VanderWeele et al. [222], who present bias formulae for interactions on both the additive and multiplicative scales.

As a complement to finding a corrected point estimate for a set of parameters describing the confounder relationship, MacLehose et al. [223] reviewed a method for deterministic nonparametric bounds for the causal effect identified by the potential outcomes model and implemented this through linear programming, a procedure more widely used in operations research and econometrics.

The general regression-based formulation developed by D. Lin et al. [224], which could accommodate censored survival times as well as continuously distributed unmeasured confounders, was the basis for the study by N. Lin et al [225]. This provided a general framework for characterising the contribution to bias from missing covariates and censoring in the Cox model. As a special case of this framework, the method of D. Lin et al. was found to perform less precisely as the magnitude of bias increases. The work by N. Lin et al. would inform later work, presenting the Pairwise framework [192], which also included a test for the presence of confounding.

Our review also picked up studies that sought to address the problem of uncertainty over the distribution of externally adjusted confounders. Steenland and Greenland [226] present Bayesian sensitivity analysis (BSA), and the analogous Monte Carlo sensitivity analysis (MCSA) as methods for acknowledging both uncertainty from random sampling and confounding bias in a single interval estimate. The BSA approach was adapted by De Vocht et al. [227] under the full Bayesian framework to adjust for smoking history as the residual confounder in a study of the risk of lung

cancer in a cohort study of workers within the European asphalt industry. With the advantage of extra processing power afforded by advances in computer technology, the authors were able to specify the confounder prevalences without approximation from the more realistic Dirichlet process as originally suggested by Steenland and Greenland. Later Carrao et al. [228] also expanded on the SA method by Steenland and Greenland [226] through their study of mono- and combination therapies and cardiovascular disease, using health records. Here the exposure-confounder association was evaluated using an external dataset, reflecting this random uncertainty through Monte Carlo sampling for three different hypothesised outcome-confounder associations.

The review also captured two papers developing BSA from the same authors. McCandless et al. [229] used both simulated data and health records in their study into treatment for heart failure to demonstrate the advantage of acknowledging uncertainty over the prior distribution of an unmeasured confounder. In a following paper [230], using the same illustrative example, the authors demonstrated that uncertainty over a single binary unmeasured confounder could be reduced using hierarchical prior distributions in BSA, by assuming the confounder originated from the same distribution as the measured confounders. Subsequently, Gustafson et al. [231] developed a simple prior distribution for a SA, requiring only a small number of hyperparameters, to model both poorly measured and unmeasured confounding. The intention was that this simplified approach, referred to as the simplified Bayesian sensitivity analysis, could be more easily adopted in practice.

A propos of poorly measured confounding, Brunelli et al. [232] studied a problem specific to retrospective cohort data, such as EHRs. Baseline data and patient histories are often gathered from information from the prior period preceding the index date of the study, here referred to as the “look-back window”. The authors, examining the sensitivity of results to the length of look-back window, found that all-available information resulted in less bias with lower mean square error compared to a fixed-interval window.

In their seminal work on PSC as a QE method, Stürmer et al [233] also presented this as a sensitivity analysis. Besides importing information about confounding from an external source, this also confers the advantage reducing the dimensionality of

multiple confounders to a single score for adjustment of the exposure effect. The PSC approach was presented in a subsequent paper by Schneeweiss [234] captured by our literature search, which presented PSC along with MCSA, the “rule-out” approach (the level of confounder imbalance and confounder-outcome association required to account for the observed effect) and what the author dubbed the “array approach”, which estimates the response surface of the outcome risk ratio after adjustment across a constellation of confounder parameters. Further work by Lunt et al. [79] presented a framework for advancing PSC as an SA method through the use of DAGs. Testing one of its key assumptions, surrogacy, they confirmed earlier work [78] that PSC was unbiased conditional on the effects of the unmeasured confounders being in the same direction as that of the observed.

The problem of dimensionality in representing an array of potential unmeasured confounders was explored in the study by Li et al. [235], which proposed the propensity score as the sensitivity function of unmeasured confounders, developed by Robins et al. [236] and Brumback et al. [237] in their SA for inverse-probability weighted estimators. This approach also demonstrated some robustness against the misspecification of the functional form of the propensity score. Again through inverse-probability weighting, but more pertinent to modelling repeated outcomes with time-dependent confounding from longitudinally observed data, Ko et al. [238] performed an SA to a plausible range of the selection-bias parameter in the marginal structural model, as developed by Robins [239] and Hernán [240].

The concept of perturbation [80], already discussed as an adjustment method in the review of QE designs, intuitively seems better suited as an SA, as it tests for unmeasured confounding through weak associations of 100’s, if not 1000’s of observed variables. This data mining approach should in theory be applicable to the big data of EHRs, but its implementation may be impeded by the method’s assumptions and the need for so many observed variables, presumably just a subset of which would comprise enough confounders to sufficiently adjust for bias.

In contrast to perturbation testing, which relies on the availability of non-confounding covariates associated with confounders, the study by the Richardson et al. [125] applied the negative control outcome (NCO) method, which takes a more parsimonious approach to utilising internal information from the dataset (as opposed

to external adjustment or external calibration data). Their study was based on earlier work by Lipsitch et al. [126], not captured in the literature search, which proposes finding another outcome unrelated to the outcome of interest and only causally related to the exposure through the confounders of exposure and outcome. This way, confounding bias may be detected through an effect of the exposure on the NCO, assumed to be through the unmeasured confounder, when conditioned on the measured confounder.

*Table 3:* Summary of studies on sensitivity analysis, returned by the literature search.

| Authors | Year | Title | Summary |
|---|---|---|---|
| Margolis, D. J.<br>Berlin, J. A.<br>Strom, B. L.<br>Berlin, J. A.<br>Strom, B. L. | 1999 | A comparison of sensitivity analyses of the effect of wound duration on wound healing | Effect of dichotomised wound duration on failure to heal in chronic leg ulcers comparing two approaches to SA, one of Rosenbaum [219] & one of Greenland [217], of estimates to unmeasured confounders. An important methodological statement on a complementary approach in using 2 SA methods together |
| Du, X. L.<br>Key, C. R.<br>Osborne, C.<br>Mahnken, J. D.<br>Goodwin, J. S.<br>Key, C. R.<br>Osborne, C.<br>Mahnken, J. D.<br>Goodwin, J. S. | 2003 | Discrepancy between consensus recommendations and actual community use of adjuvant chemotherapy in women with breast cancer | Study investigated whether chemotherapy varies with age in women, and tested sensitivity of results to unknown confounders using method by Greenland [217]: This was expanded for eight-level exposure groups of age, and tested sensitivity to a confounder dichotomised around different age cut points. |
| Ko, H.<br>Hogan, J. W.<br>Mayer, K. H. | 2003 | Estimating causal treatment effects from longitudinal HIV natural history studies using marginal structural models | Investigation of effect of highly active antiretroviral therapy regimens on CD4 cell counts in 871 HIV-infected women recruited for the HIV Epidemiology Research Study, using Marginal Structural Models / G-estimation. It implements (in what is posited as the first example) of Robin's approach [239] to SA by estimating the bias as the difference between the counterfactual means, given observed confounders, but here for a plausible range of selection bias in prescribing. |

| | | | |
|---|---|---|---|
| Steenland, K.<br>Greenland, S. | 2004 | Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer | Authors discuss the advantages of Monte Carlo (MC) sensitivity analysis (MCSA). Ordinary sensitivity analysis only allows postulation of a point estimate of bias, usually when only the direction of bias can be ascertained with any certainty. Bayesian and MC approaches can generate a range of plausible adjusted effect estimates for a given distribution of bias. However, while the implementation of a Bayesian sensitivity analysis (BSA) can be involved, requiring complex understanding, MCSA approximate BSA methods under certain conditions. This was demonstrated on a cohort of workers' silica exposure and lung cancer outcomes, adjusting for bias from smoking habits, the distribution for which is based on 1987 survey data. |
| MacLehose, R. F.<br>Kaufman, S.<br>Kaufman, J. S.<br>Poole, C. | 2005 | Bounding causal effects under uncontrolled confounding using counterfactuals | A nonparametric approach is presented in this study as a complement to sensitivity analysis, through linear programming for determining the bounds, or absolute limits, of the true effect of the exposure in the presence of unmeasured confounding. This was done using the observed table of observed data and counterfactuals, under realistic assumptions about the potential outcomes. Method applied to exemplar of effect of beta-blockers on mortality. |
| Stürmer, T.<br>Schneeweiss, S.<br>Avorn, J.<br>Glynn, R. J. | 2005 | Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration | Seminal paper on propensity score calibration (PSC) used to adjust for unmeasured confounding in study of effect of NSAIDs (non-steroidal anti-inflammatory drugs) on 1-y mortality in elderly, through use of a calibration dataset to complete adjustment for unmeasured confounders in main error-prone dataset. Here the main data is the Medicaid register of adults aged >= 65y with hospitalisations 1995-97, calibrated with Medicare Current Beneficiary Survey. PSC was proposed as a SA until method limitations & validity assessed |

| | | | |
|---|---|---|---|
| Malay, D. S.<br>Margolis, D. J.<br>Hoffstad, O. J.<br>Bellamy, S.<br>Margolis, D. J.<br>Hoffstad, O. J.<br>Bellamy, S. | 2006 | The Incidence and Risks of Failure to Heal After Lower Extremity Amputation for the Treatment of Diabetic Neuropathic Foot Ulcer | The study comprised an exploratory analysis of risk factors and treatments in failure to heal of extremity amputation in the diabetic foot. The sensitivity of the results to unmeasured confounders was tested by Greenland's SA [217]: Estimates were concluded to be robust to confounding |
| Schneeweiss, S. | 2006 | Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics | This study presents a tutorial-like review of SA methods the selection of which will be dependent on the availability of information about the (binary) confounders (affecting binary outcomes):<br>The array approach estimates adjusted relative risks (RR) & bias across parameters for RR of the confounder-outcome association and confounder prevalence in each group<br>Rule-out: equation allows estimation of RR of the confounder-disease association and confounder prevalence in groups to move the apparent RR to the null.<br>External adjustment: This allows a more straight-forward estimate of RR confounder-disease association when the confounder prevalence is known from an external validation data<br>External adjustment for multiple confounders: describes PSC.<br>Simulation via MCSA<br>The array approach estimates adjusted relative risks (RR) & bias across parameters for RR of the confounder-outcome association and confounder prevalence in each group<br>Rule-out: equation allows estimation of RR<br>External adjust<br>External adjustment for multiple confounders: describes PSC.<br>Simulation via MCSA |

| | | | |
|---|---|---|---|
| McCandless, L. C.<br>Gustafson, Paul<br>Adrian, Levy | 2007 | Bayesian sensitivity analysis for unmeasured confounding in observational studies | Using real and simulated data, the study looks at the consequences of non-identifiability of the prior distribution and the effect of prior misspecification on the estimated interval of Bayesian SA. By acknowledging uncertainty about unmeasured confounding, the authors demonstrated that credible Intervals achieve approximately nominal coverage probability if prior distribution is similar to sampling distribution of the model parameters. According to average coverage probability, it is better to acknowledge uncertainty about unmeasured confounders, even if model for BSA is unidentifiable. |
| Arah, O. A.<br>Chiba, Y.<br>Greenland, S. | 2008 | Bias Formulas for External Adjustment and Sensitivity Analysis of Unmeasured Confounders | The authors review and present a simple set of bias expressions for SA of effects on different scales (risk difference, risk ratio, odds ratio), and demonstrate how this can be extended to polytomous confounders, exposures and outcomes. |
| Cabral, M. D.<br>Luiz, R. R. | 2008 | Use of sensitivity analysis to assess the effects on anti-hepatitis A virus antibodies of access to household water supply | The study used two approaches to SA, citing Rosenbaum [219] and then Greenland's [217] external adjustment to evaluate the plausibility of results from investigation into the effects of access to household water on hepatitis A prevalence, and found the odds ratio between confounder and outcome would need to be >=4 to explain the observed effect. |

| | | | |
|---|---|---|---|
| McCandless, L. C.<br>Gustafson, P.<br>Levy, A. R.<br>Gustafson, P.<br>Levy, A. R. | 2008 | A sensitivity analysis using information about measured confounders yielded improved uncertainty assessments for unmeasured confounding | Using external data on unmeasured confounding might not narrow the uncertainty over confounding. However, the authors proposed using existing associations between the measured confounders and prognosis to better inform the interval estimates of the treatment effect by assuming a similarity of the unmeasured confounders and treating the coefficients of measured and unmeasured confounders as random samples from a normal distribution. Shrinkage estimates can then take account of the similarities in a Bayesian treatment of hierarchical models. |
| De Vocht, F.<br>Kromhout, H.<br>Ferro, G.<br>Boffetta, P.<br>Burstyn, I.<br>Kromhout, H.<br>Ferro, G.<br>Boffetta, P.<br>Burstyn, I. | 2009 | Bayesian modelling of lung cancer risk and bitumen fume exposure adjusted for unmeasured confounding by smoking | Authors adopted & expanded Steenland & Greenland's [226] BSA to assess the effect of smoking, identified as the unmeasured confounder in the risk of lung cancer to bitumen exposure<br>The study assumed informative priors based on a Dirichlet distribution. The posterior distribution was generated by Gibb's sampling using Metropolis MCMC. The association of lung cancer with bitumen exposure was still supported after SA, but with reduced certainty.The study assumed informative priors based on a Dirichlet distribution. The posterior distribution was generated by Gibb's sampling using Metropolis MCMC. The association of lung cancer with bitumen exposure was still supported after SA, but with reduced certainty. |

| | | | |
|---|---|---|---|
| Fung, T. T.<br>Van Dam, R. M.<br>Hankinson, S. E.<br>Stampfer, M.<br>Willett, W. C.<br>Hu, F. B.<br>Van Dam, R. M.<br>Hankinson, S. E.<br>Stampfer, M.<br>Willett, W. C.<br>Hu, F. B. | 2010 | Low-carbohydrate diets and all-cause and cause-specific mortality: Two cohort studies | In their study on the effects of low-carbohydrate diets (LCD) on cause-specific mortality using two cohorts of data, the authors cited Lin et al [224] in their SA, and found that the confounder imbalance would unlikely be strong enough to entirely explain the apparent effect of higher mortality from animal-sourced LCDs and lower mortality from plant-based LCDs. |
| Gustafson, P.<br>McCandless, L. C.<br>Levy, A. R.<br>Richardson, S. | 2010 | Simplified Bayesian Sensitivity Analysis for Mismeasured and Unobserved Confounders | Simple Bayesian Sensitivity Analysis (SBSA) is developed from the authors' previous work, and aims to incorporate poorly measured confounders, as well as unmeasured confounders, that are realistically likely to be available in most data. Focus is on simplifying the specification of the prior and hyperparameters involved |

| | | | |
|---|---|---|---|
| Li, L.<br>Shen, C.<br>Wu, A. C.<br>Li, X.<br>Shen, C.<br>Wu, A. C.<br>Li, X. | 2011 | Propensity score-based sensitivity analysis method for uncontrolled confounding | The authors propose a 1-dimensional function of the propensity score, referred to as the sensitivity function, to quantify bias due to hidden confounders, and as an alternative SA for IPW estimators by Robins et al. [236] and Brumback et al [241]. The SF was used to correct IPW estimators. Advantage of method is it reduces dimensions of confounders and polynomial forms of PS function can be reasonably approximated by lower order polynomials. The method is demonstrated on study of medication frequency in asthmatics. |
| Corrao, G.<br>Nicotra, F.<br>Parodi, A.<br>Zambon, A.<br>Soranna, D.<br>Heiman, F.<br>Merlino, L.<br>Mancia, G.<br>Nicotra, F.<br>Parodi, A.<br>Zambon, A.<br>Soranna, D.<br>Heiman, F.<br>Merlino, L.<br>Mancia, G. | 2012 | External adjustment for unmeasured confounders improved drug-outcome association estimates based on health care utilization data | The authors applied SA to a nested case-control study of the effect of different drug regimens of mono and combination antihypertensive therapies on cardiovascular outcomes. Steenland & Greenland's Monte Carlo SA [226] was also applied to acknowledge the uncertainty over the distribution of bias and to incorporate external information about the confounders from other data sources. The results explained the apparent different in risk between two drugs administered extemporaneously and in combination, but not between combination and mono-therapies. |

| | | | |
|---|---|---|---|
| Lunt, M.<br>Glynn, R. J.<br>Rothman, K. J.<br>Avorn, J.<br>Sturmer, T. | 2012 | Propensity score calibration in the absence of surrogacy | Authors present a framework by way of directed acyclic graphs for using PS calibration to adjust for confounding through an external, cross-sectional validation dataset in the simple scenario of one measured and one unmeasured confounder. A formula can be utilised to predict the presence and magnitude of bias in PSC. Under the assumption of independence between the observed & unobserved confounders, only one parameter needs to be substituted in the bias formula. Simulated cohorts were analysed and the effectiveness of NSAIDs on survival times of older adults estimated, using influenza vaccination status and age as gold-standard and error-prone PS's respectively |
| Vanderweele, T. J.<br>Mukherjee, B.<br>Chen, J. | 2012 | Sensitivity analysis for interactions under unmeasured confounding | This technique was developed for assessing the sensitivity of interaction analyses to unmeasured confounding, with presentation of the bias formulas. This was demonstrated using data on the interaction between passive smoking and glutathione S-transferase M1 (GSTM1) on the risk for 106 lung cancer cases among non-smokers. |
| Brunelli, S. M.<br>Gagne, J. J.<br>Huybrechts, K. F.<br>Wang, S. V.<br>Patrick, A. R.<br>Rothman, K. J.<br>Seeger, J. D. | 2013 | Estimation using all available covariate information versus a fixed look-back window for dichotomous covariates | SA of the differential effects of missing information on covariates from medical records, which results in unmeasured confounding bias. Simulations compare the benefits of fixing the window of time for collecting the information from all subjects to collecting all available information as determined by the subjects' historic contact and registration with the medical system. The simulation derives estimated bias for a range of probabilities for binary covariates, exposures and outcomes, for which recorded healthcare utilisation is a probability dependent on a binary frailty variable. The conclusion is that less bias is introduced when all available information is introduced rather than over a common fixed window. |

| | | | |
|---|---|---|---|
| Lin, N. X.<br>Logan, S.<br>Henley, W. E. | 2013 | Bias and sensitivity analysis when estimating treatment effects from the cox model with omitted covariates. | This study extends the formula for bias due to confounding in the Cox model by Lin D et al [224] to a general framework for other potential sources of bias, applying this to randomised and observational studies and simulated data. It concludes that the results from a Cox model are biased by missing covariates, even if those covariates are balanced, and that censor bias is maximised at 50% censoring. |
| Lee, W. C. | 2014 | Detecting and correcting the bias of unmeasured factors using perturbation analysis: a data-mining approach | Presentation of proposed method to test for unmeasured confounding (perturbation test) and to adjust for it (perturbation adjustment) where very large data exist through a data mining approach by accounting for the weak associations with any unmeasured confounders of multiple perturbation variables (PV), possibly 100's, if not 1000's. However the large number of PVs may exceed the available df's, so clustering of PV levels may be necessary. Longitudinal data may mitigate against incorrect adjustment of collider. Inadvertent adjustment of a measured confounder, however, exacerbates the bias, which may be eventually attenuated with many more PVs. More work needed to explore continuous PVs. |

| | | | |
|---|---|---|---|
| Richardson, D. B.<br>Laurier, D.<br>Schubauer-Berigan, M. K.<br>Tchetgen, E. T.<br>Cole, S. R. | 2014 | Assessment and indirect adjustment for confounding by smoking in cohort studies using relative hazards models | This cites earlier work on negative control outcomes (NCO) [126].The study of lung cancer risks in occupational cohort studies, usually with unverifiable confounding effects from unmeasured smoking status, is used as the exemplar in this presentation of the use of an NCO in the SA and adjustment for unmeasured confounders such as smoking status. Here the measured confounders and the unmeasured confounders are common causes of the NCO. Adjustment for NCOs are an improvement on ordinary SAs, which make unverifiable assumptions about the prevalence of smoking among exposure groups, but the method imposes other unverifiable assumptions including that the exposure does not cause the NCO. Here COPD was used as the NCO in the study of the occupational risk of lung cancer in Colorado miners and French nuclear workers. |

## 3.4 Discussion

SA often relies on assessing whether a hypothesis about the size of the confounding effect required to explain the observed effect of exposure is credible. However, an implausibly large confounding effect does not completely rule out bias, the size of which may still be clinically meaningful. In this way, BSA and MCSA have proved to be useful in acknowledging the uncertainty over the distribution of confounding bias. External data sources may be also useful in yielding more information about confounders as in the propensity score calibration methods. Many confounders are likely to be associated so basing hypotheses about unmeasured confounders on those observed will likely be informative. Propensity scores may also be useful as a sensitivity function, reducing the dimensions of many different confounders into a single summary.

In contrast to external adjustments, the method of negative control outcomes stands out as one that makes use of available information within the same dataset. Another, source for information on unmeasured confounding may be found within the patient histories of the dataset. Indeed, many of the before-and-after methods reviewed in chapter 2 rely on the longitudinal information from patient histories, as well as the longitudinal adjustment for measured confounders using weighted methods (broadly known as G-methods). In this respect, the study by Brunelli et al is highly informative in determining the look-back window for patient histories.

Where the ability to capture information about confounders is in doubt, many of the QE methods may be regarded as SAs themselves, as has been suggested for the method of propensity score calibration. As will be discussed in the next chapter, the prior period of the PERR method and the control arm of the Pairwise method may offer a window onto the direction and size of confounding. In a comparison of the two methods, the performance of the PERR and Pairwise methods will be greatly informed in chapter 4 by the work of Lin, Logan and Henley [225].

## 3.5 Conclusion

Sensitivity analyses aim to indirectly characterise the relationship between exposure, outcome and unknown confounder(s). From the reviewed literature, the approach taken to SA will be informed by the characteristics describing assumptions about the unmeasured confounder:

- Hypothesised confounders vs. information on confounders from external data sources vs. indirect internal information on confounders
- Single or multiple confounders vs. single dimensional summary
- Point estimates for SA vs. methods to integrate uncertainty over unmeasured confounders
- Modelling vs explicit formulation for SA

Hypothesising a range of confounders, or deliberating over the size of confounding required to move the observed effect to the null may be akin to plucking numbers out of the air, unless there is some reference. If knowledge about the unobserved confounders can be gleaned from another dataset, then these may serve as a reference. If the information on these is complete, then a calibration may be possible using PSC, although this may prove to be quite rare in practice. However, in a dataset sufficiently rich in information, finding an NCO would obviate the need for an external dataset, as might patient histories, if the degree of confounding bias may be gleaned from the longitudinal information. Many of the methods, such as NCO and PSC, may be characterised as an empirical, model-based approach. However, explicit formulae exist for re-calculating the odds ratio or relative risk adjusted for a hypothesised confounder. Here, the effect of a single confounder may be represented, but in reality, confounding bias is likely to be from multiple sources. Propensity scores and dimension reducing functions of multiple confounding effects may offer a means of modelling realistic scenarios. However many confounders may be unobserved, the resulting bias can only be unidimensional. Yet, uncertainty will inevitably exist over the true degree of bias. Here, this can be accommodated as a hypothesised distribution of the bias through Bayesian and Monte Carlo SA methods. The need for positing a distribution for the confounding bias may be

obviated, however, by the availability of indirect information about the bias within the same dataset.

# Chapter 4 - The PERR and Pairwise methods: review and implementation

## 4.1 Development of the PERR and Pairwise methods

Attempts to replicate trials in a 2008 study using the observational data of EHRs prompted the authors, Tannen, Weiner and Xie, to develop a new method to address unmeasured confounding [242,243]. The prior event rate ratio (PERR) method was an intuitive before-and-after approach, analogous to differences-in-differences, that could be applied to both survival and rate data. Although, quasi-experimental methods are available for continuous and binary outcomes, there are few, if any, methods suitable for survival data. While rate and survival data are two possible measurements of the outcomes generated by the same process, the attraction of using survival data is that it is possible to analyse the data without modelling the underlying distribution of survival times by applying the popular Cox regression model (also known as the proportional hazards model) [244]. Extensive literature exists on the theory and application of the Cox model. Briefly, the effect of an exposure or other variables of interest may be estimated from the maximum likelihood of the joint probability of the events as a function of the variables' coefficients for individuals ordered by their survival times observed over a prescribed period. As the method relies only on the rank of the ordered survival times rather than the length of the survival times themselves, this is also known as the partial likelihood, expressed as the likelihood of the exponent of the linear predictor of a vector of coefficients, β, for ***x*** covariates, observed at the $k^{th}$ ordered event of *i* times:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{exp(\boldsymbol{\beta}.\boldsymbol{x}_{(i)})}{\sum_{l \in \Re(t_{(i)}} exp(\boldsymbol{\beta}.\boldsymbol{x_l})}$$

*Equation 1*

where the denominator is the set of individuals at risk at the $i^{th}$ ordered event as stated in Cox [244]. The times may terminate with an event, in which case the time of that individual is present in both the numerator of the probability and in the denominator as part of the risk set of all survival times lasting up to that point in time.

The times that do not terminate in an event are censored and are only in the denominator of the joint probability function as part of the risk set for all survival times of individuals at risk at the time of the event in the numerator. This approach avoids direct estimation of the underlying baseline hazard function.

Outside actuarial sciences and process monitoring in manufacturing, survival data are perhaps mostly commonly encountered in medicine. Therefore the arrival of the PERR method, was key to the development of more inferential studies using routinely collected data at a time of growing interest in using EHRs for health research [245].

### 4.1.1 The PERR method

The PERR method proposes that if the hazard (or rate) ratio of the exposure in a period of study is biased by confounding, this may be adjusted by the hazard (or rate) ratio between the two exposed groups from a period prior to the exposure (Figure 4). This, therefore, demands an exposure (treatment)-free period prior to the period under study, and assumes that the hazard (or rate) ratio from the exposure-free prior period is a measure of the pre-existing confounding bias between the two exposure groups. While the PERR method offers a possible route for addressing unmeasured confounding, like all quasi-experimental methods, it is appropriate for certain types of data, and its reliability rests on meeting certain assumptions (see below). Since information on the outcome from an exposure-free prior period is key to the correction for confounding bias, then the outcome needs to be repeatable. Implicit in the assumption of the prior period offering a correction for confounding bias, is that the bias is constant from one period to the next. Upon meeting these conditions then the PERR estimate is provided by either the hazard from a survival model or the incidence rate from a suitable model, such as the Poisson, in the study period divided by the corresponding estimate from the prior period. In this way, the estimate from the prior period should provide an estimate of the magnitude and direction of confounding that may be present in the study period, if the assumptions are correct.

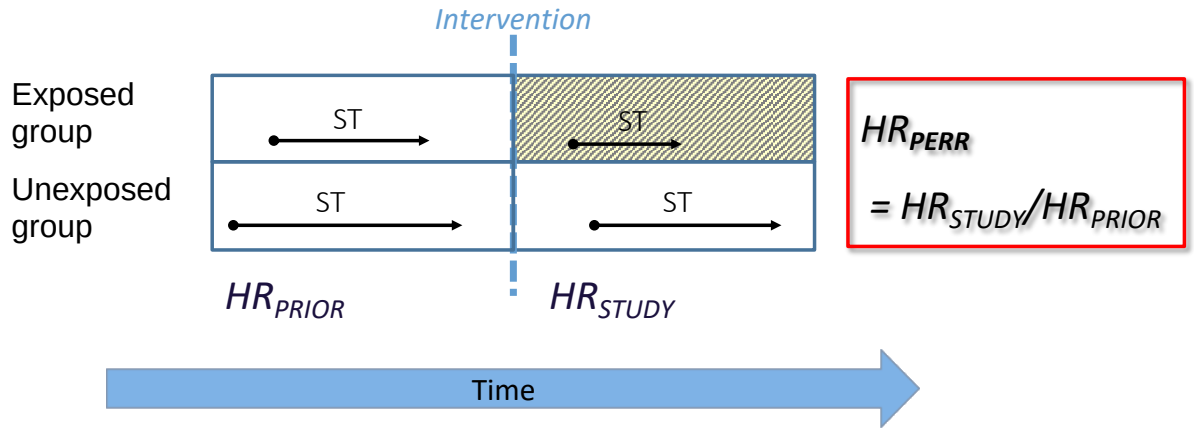


*Figure 4:* Diagram of the adjustment made across two periods in the prior event rate ratio (PERR) method. The hazard ratio (HR) of intervention is calculated from survival times (ST) from the study period and adjusted with that from the exposure (treatment)-free prior period. The shaded cell denotes the exposure arm in the study period. The PERR adjustment can also be applied to rate data.

### 4.1.2 The PERR-Alt or Pairwise method

In a follow-up to the seminal paper, an alternative formulation of the PERR method, the PERR-ALT method, was first presented by Yu et al [246]. In contrast to adjusting the Cox estimate from the study period with that from the prior, the PERR-Alt method is based on a paired Cox analysis, in which an adjustment for a period effect is made within each subject. Under the large sample approximation, the difference between the PERR and PERR-ALT method is clear: The adjustment for confounding using the PERR-ALT method is made within each of two exposure groups. The hazard ratio of treatment effectiveness is then the ratio of the hazard ratios from each adjusted exposure group. Conversely, the PERR method first finds the hazard ratio of the exposure effect in the study and prior period and then adjusts the hazard ratio of exposure in the study period with that from the prior period. Results from a limited array of simulations in the study by Yu et al. demonstrated no bias in the PERR-ALT and little bias in the PERR method relative to exposure effect assuming equal confounding effects in both study intervals. Further work was also discussed, but not presented for the effects of prior events on those in the study period.

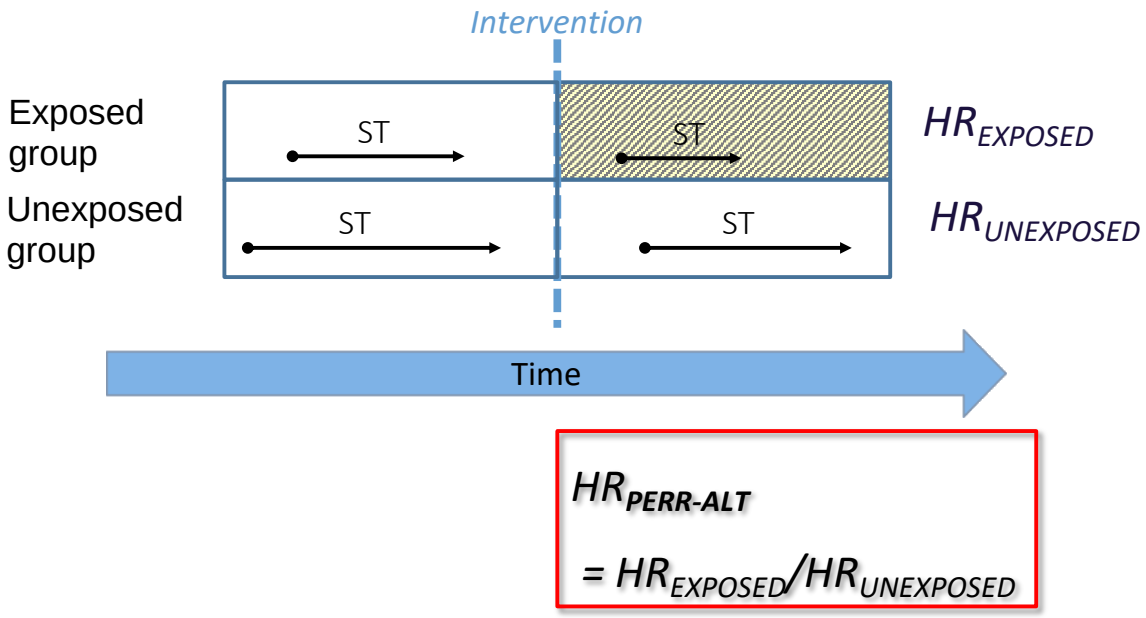


*Figure 5:* Diagram of the adjustment made within each exposure arm between two periods using the PERR-ALT approach. The hazard ratio (HR) of the exposure in the study period relative to the prior period is calculated from the survival times (ST) for the exposed group and then adjusted with the analogous ratio for the unexposed group. The shaded cell denotes the exposure in the study period. The PERR-ALT adjustment can also be applied to rate data.

Further results from simulations were presented in another method paper by Uddin et al.[247], under a variety of scenarios, which will be discussed in more detail below. Simulations also examined the relative performance of the PERR and PERR-ALT methods together in the method paper by Lin and Henley that presented a generalised formula for the Pairwise likelihood method [192], of which the PERR-ALT method is a special case. The formula extended the Pairwise method to accommodate time-varying covariates as well as multiple periods, and demonstrated superiority in dealing with time-varying baseline hazards, although a greater sensitivity to confounder-treatment interactions. Furthermore, the authors demonstrated that the problem of bias will stubbornly afflict estimates from the application of the PERR method. This is due to problems of bias with the underlying Cox models in the presence of missing covariates as demonstrated in earlier work [225]. The problem of bias in the PERR estimates was shown to be exacerbated in the presence of censoring. Only when the missing covariate is a confounder does the bias in the Cox model reduce under the PERR method.

## 4.2 Clarification of types of confounding

In order to examine the assumptions around the two methods, it is first necessary to clarify some the terminology around confounding. Two terms for confounding bias are frequently encountered, sometimes within the same study. These are "confounding by indication" and "selection bias". It would be interesting, but, perhaps not especially worthwhile given the work involved, to conduct a systematic review to study consistency of usage of these terms. However, there can only be one effect of confounding: bias. The two terms simply describe the mechanism, by which confounding might arise, and it is important to clarify the motivation for using either "confounding by indication" or "selection bias" instead of "confounding bias". As established in the introduction chapter, the first condition for a confounder in Statistics is that it must be any other variable, other than the variable(s) of interest, which may affect the outcome(s) of interest. In medicine and health, this is likely to be referred to as a prognostic variable. As variables of interest are often binary exposure or treatment variables, then any potential prognostic variables will confound the exposure variable if these are imbalanced between exposure groups. Patients may be selected for a particular treatment based on prognostic factors, or characteristics associated with these. In this way, the second condition for a confounder is met by association of the prognostic variable either through imbalance between exposure groups or through selection for treatment based on it. In the UK, the patients may seem to fare badly and be worse off under particular treatments, if these are prioritised for those most in need. In countries like the USA, access to healthcare largely depends on income. Better health may be seen to be enjoyed by those with a higher socioeconomic status, and so this confounding may lead to a healthy patient bias. In both cases, some form of *selection bias* has occurred.

*Confounding by indication* is subtly different. A risk of bias still exists from what is a form of selection, but here "by indication" is taken to mean selection on the outcome or associated event, which subsequent treatment is intended address. This is an important distinction in the context of the PERR and Pairwise methods applied to longitudinal EHR data, as the definition implies that confounding-by-indication can only occur when an outcome is repeatable – an essential feature of data required for the application of the PERR and Pairwise methods. In the case of presentation of a

disease, “by indication” would therefore be the prescribing of a treatment intended to prevent future recurrence of the disease.

## 4.3 Assumptions of the PERR and Pairwise methods

The assumptions have variously been presented and tested through simulation in the seminal studies discussed above. For clarity, these are consolidated and discussed below, and each simulation study critiqued where necessary:

1. *The same net effect of confounding in the prior as in the study*

This refers to the condition of there being no difference in the underlying bias between the study and prior periods. This does not explicitly require the same confounders to be present in both periods, but rather the same effect is exerted. In the simulation study of Uddin et al.[247], different symbols were used to represent the confounders ($C_{11}$ for the prior and $C_{12}$ for the study period) in order to simulate different period-specific effects of confounding. The implication from the arrow from $C_{11}$ to $C_{12}$, that confounders in one period “cause” those in the other may be contentious, especially where these are asserted to be equal. Such an issue could be treated as further evidence of problems in using DAGs to illustrate observed epidemiological relationships [35]. Theoretically, a different set of confounders for each period could by chance exert the same effect with one causing the other, which may warrant different symbols, although in practice it is difficult to conceive a situation where this would happen to be the case. In practice, it would be necessary to accept no significant difference between the periods’ confounding effects, given some non-significant differences may persist as a result of modelling instability and random variation. If the same confounders are present throughout, then their effect is assumed to be constant over the prior and study periods. In their presentation of the PERR-ALT method, Yu et al. demonstrated that the bias from the period-confounder interaction relative to the exposure effect was “well-controlled” at less than 5% when the exposure was small, and less than 10%, when the exposure effect was “moderate to large”[246]. If the confounding effect is stable then this suggests time-invariant confounding between the two periods. The property of time-invariant confounding may be reasonable when:

   a. the population characteristics are stable over time.

The underlying latent factor that broadly encapsulates prognostic factors and susceptibility to disease could be described as frailty. The precise definition of frailty and the variables that measure it may change according to the disease, but the rate of change in frailty over time may differ according to the population under study. Among the very old, or ill patients with a poor prognosis, the rate of change in some "frailty" variables may be great enough to consider restriction of the observation time if any of these are confounders. If the change in the confounding effect is too great between the periods, then the prior period will be a poor adjustment for confounding in the study period.

b. the prior and study periods are adjacent to each other:

The proximity of the start of the prior period to that of the study period depends in part on the length of the prior period. In the case of certain annual vaccinations, such as those for influenza, the proximity will also be determined by any seasonality in the treatment and follow-up. The proximity of the periods may also depend on the availability of an exposure-free sample for recruitment. If in the case of the phased-in introduction of health policies, such as vaccination, a larger cohort may be assured by looking back far enough to maximise the availability of treatment-free individuals. However, this will need to be balanced against the stability of confounders, and that of frailty as discussed in (a), as over time the relationships with any confounding variables are likely to change.

c. the prior and study periods are short:

For reasons discussed in (a), the length of the study and prior periods will impact on their proximity as noted in (b). However, longer periods will likely allow more events to observed as well as a longer recruitment period, both of which would increase the power of a study, so a balance invariably needs to be struck between these considerations and the violation of time-invariant confounding.

2. *No moderating effects of confounders on exposure or treatment*

In their presentation of the Pairwise method and evaluation of the relative performance, Lin and Henley presented results for the PERR and Pairwise methods compared against the Cox model under simulation of an interaction between the confounders and exposure effects [192]. Under parameters of unity, the PERR

adjustment was seen to perform better than the Cox model, while the Pairwise method only seemed to be less biased than the Cox model when the magnitude of the interaction effect was less than that of the main effect in this particular scenario. However, it was not clear how this might be attenuated under different parameters.

3. *Independence between outcomes*

The occurrence of events should be independent so that any event should not change the probability of any subsequent events. More specifically to the PERR and pairwise methods, events in the prior period should not directly influence those in the study period. This may be more of a problem when events are common enough to occur more than once within the same patient. However, for the scenario presented by Uddin et al. [247], the bias was relatively small compared to the conventional risk ratio. Nevertheless, this may be questionable if, say, in the case of infectious diseases, infection in a prior period conferred immunity in the study period.

4. *No causal effect of the prior outcome on exposure*

As explained in the section above, where the outcome in the prior period determines the exposure in the study period, this can be regarded as *confounding by indication*. The implication for the two-period study design analysed by the PERR method is that the effect on the outcomes from the imbalanced prognostic variables (i.e. confounders) is accompanied by a causal effect of the outcomes in the prior period on exposure. In this matter DAGs can seemingly both clarify and confuse understanding of this concept. To understand how this might arise, it is helpful to first consider *indication* (effect of prior outcome on exposure) in the absence of confounding. This was presented, but not fully explained, in scenario 3 of the study by Uddin et al (Figure 6) [247]. The scenario of a direct causal effect without any prognostic or confounding variables is arguably unrealistic in any observed biological system, yet this explains the problem of *indication*. In scenario 3, the exposure is determined by event $Y_1$, which must therefore occur entirely at random in the prior period since it has no cause or confounder. Since the causal path $X \rightarrow Y_2$ is unconfounded, then adjustment with what is a randomly generated process $Y_1$, where treatment X is wholly or partly determined by $Y_1$ would induce bias in the estimate of the effect of X on $Y_2$. Therefore the problem with confounding by indication in the PERR and Pairwise methods is that it comprises two effects, and an

adjustment is made not only for confounding, but also *indication* (Figure 7). In this way, adjusting for $Y_1$ as an indicator in the regression of X on $Y_2$ is akin to adjusting for an instrument in which X becomes the mediator rather than the cause of $Y_2$.


*Figure 6:* Recreation of figure 4a from scenario 3 in the study by Uddin et al., where the causal effect in the study period between X and Y2 is unconfounded, but X is indicated by prior event Y1

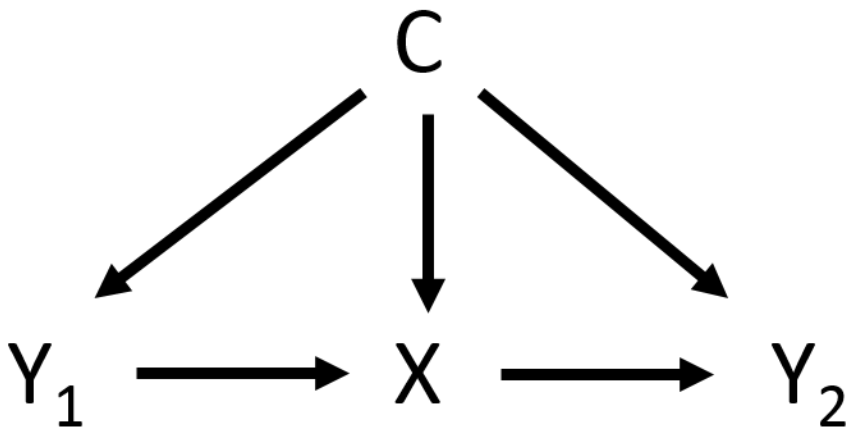


*Figure 7:* Representation of figure 4c from scenario 3 in the study by Uddin et al., where the causal effect in the study period between X and $Y_2$ is confounded by C, but X is indicated by prior event $Y_1$, which is also caused by confounders C.

The idealised relationship for the application of PERR is represented in Figure 8, which shows the prior outcome $Y_1$ and the causal effect of X on $Y_2$ confounded by C. In many clinical scenarios, one could expect the actual relationships to be somewhere between this idealised scenario and that with indication in Figure 7. An example could be where patients receive an invitation for routine vaccination based on age and irrespective of health status, but following an illness close to their appointment, some patients, who otherwise would ignore the invitation, might elect to be vaccinated.  This demonstrates the confusion that DAGs may cause, as they can

falsely represent a dichotomisation of a relationship that may exist on a continuous scale between $Y_1$ acting as a proxy for confounders C and $Y_1$ as an instrument for X. Certainly, the strength of $Y_1$ as a proxy for confounders C would be stronger in some clinical scenarios, for instance, in a study of the side-effects of a treatment. Uddin et al. showed under simulation that the logarithm of the hazard ratio moves linearly away from the point of no bias with an increasing effect of indication.

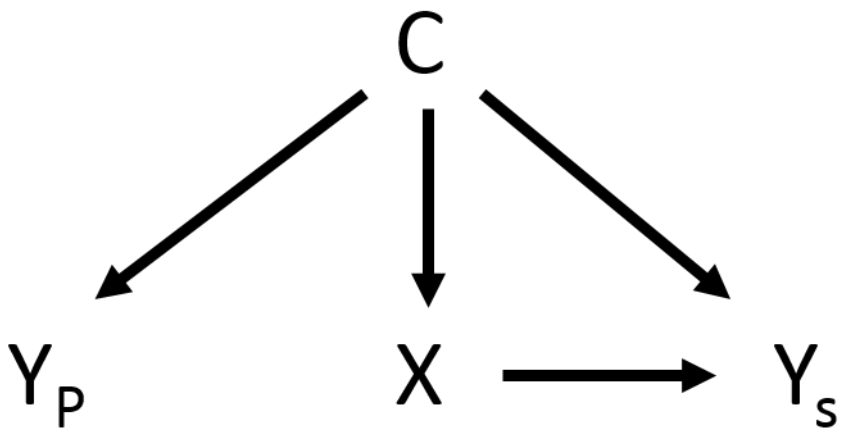


*Figure 8:* PERR is applied ideally to a causal relationship between exposure, X, and study outcome, $Y_S$, confounded by C. Prior outcome, $Y_P$, is also predicted by confounders, C, but crucially there is no indication between $Y_p$ and X.

The simulation of Lin and Henley showed a more complicated relationship for a different range of settings[192]. Overall the PERR estimates performed better than those from those from the Cox model, although for a wider range of values the Pairwise estimates were affected the least by bias.

5. *No differential effect of dropout or censoring*

The condition for censoring and dropout in applying the PERR and Pairwise methods depends on which of the two methods is used. Censoring is a potential problem in survival analysis, if this is deemed to be informative and imbalanced between exposure levels. Censoring may include death if this is not an outcome, although due to the unrepeatable condition of death, this would not be a valid outcome under the PERR approach in any case. The potential sensitivity of the survival model to informative censoring may be evaluated by inspecting the distribution of censored times relative to event times and by comparing the results from reassigning the

censors as events with those from assigning the longest times to censored times[248]. Inverse probability of censor weighing may be possible if there are a sufficient number of censored times. Irrespective of informativeness, however, the problem posed by dropout/mortality is that this is an unrepeatable condition in the prior period, as patients are necessarily alive up to the point of recruitment in the study period. If dropout is imbalanced between treatment groups, as simulated in Uddin et al. and Lin and Henley, then this is likely to be commensurate with an imbalance in the associated confounders, creating a problem akin to the violation of assumption 1 above: that the net effect of confounding should be equal in both periods, or at least similar enough to lead to no significant practical difference. In the simulation by Uddin et al., it was not entirely clear how imbalanced was mortality between exposure groups, but the results suggested a downward bias with increasing mortality rates. They also reported that excluding mortality in the study period numerator of the PERR method was more biased than the PERR that included all patients. Lin and Henley dubbed the difference in mortality between periods, and the potential resulting confounder imbalance, as the *differential case fatality (DCF)*, in reference to an abstract by Gallagher [249] (there was no subsequent peer-reviewed, published study, so this was not included in the method review). Their simulation confirmed that DCF arises from imbalance in confounders, but also demonstrated that DCF can decrease as well as increase the bias of the PERR method. Interestingly, they also showed that in the extreme case of all patients at one level of a binary confounder dying before the study would result in no bias. Crucially, the results from the Pairwise method were also unbiased, since the comparisons are always paired within each patient. Reflecting the concern over imbalanced censoring, Lin and Henley's steps for detect confounding in the prior period of the PERR method took account of possible bias from imbalanced censoring. They first proposed a test of significance between exposure groups having swapped the censor and event indicators, following a similar method proposed by Collett for testing the imbalance in censoring [248]. However, censoring in the prior period is unlikely to be anything other than administrative censoring of patients, who have reached the end of the prior period without an event, and so must necessarily be alive for the study period, for which they were recruited.

## 4.4 PERR and Pairwise methods in the context of general approaches to confounding

In chapter two a variety of methods were reviewed, which leveraged the longitudinal information of EHRs to adjust for unmeasured confounding. Many of these relied on longitudinal information from an exposure-free period prior to the period of study itself. While this type of before-and-after adjustment could be described as an "approach" or method, it also intrinsically determines the "shape" and features of the data and the way, in which these are analysed. Hence, study design and the method used are not mutually exclusive for many of the approaches. In this way, it would be misleading to consider the PERR and Pairwise methods in isolation of all other potential strategies for minimising confounding bias.

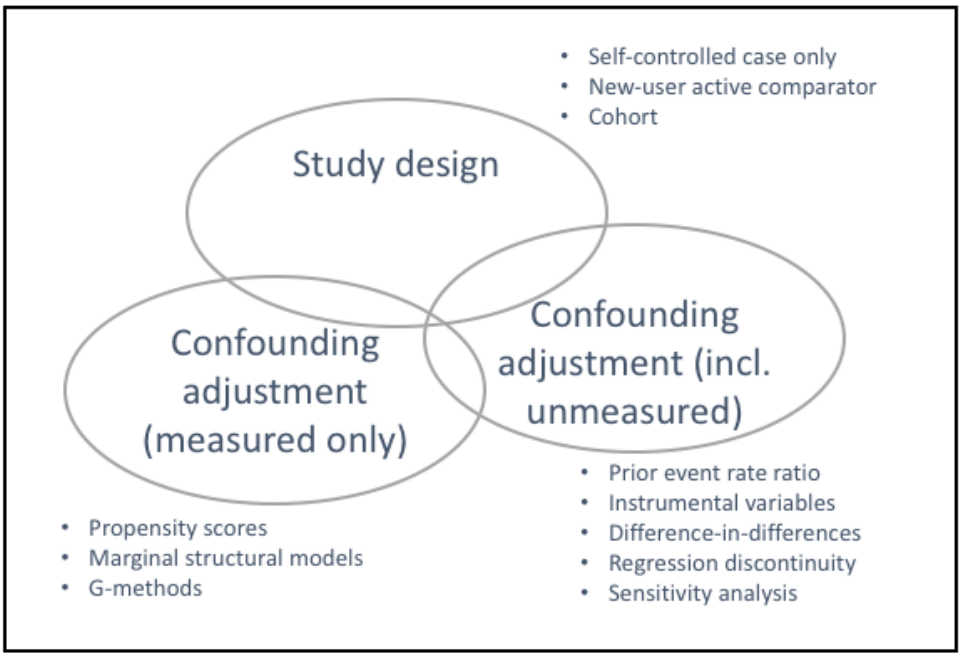


*Figure 9:* Venn diagram of possible approaches to reduce or mitigate for confounding bias with examples (not exhaustive) of each.

In practice, it is useful to consider how different approaches may overlap and be combined to minimise the risk of bias due to confounding (Figure 9). For example, some form of restriction to a particular sub-group or within strata of the population may control for confounding by aiming to reduce the difference in prognostic characteristics between exposure groups and may be deployed as a form of

sensitivity analysis of the reliability of the main unrestricted result [250]. This is a common approach, and recently there has been a growing interest in the test-negative case-control design particularly for monitoring annual vaccine effectiveness [251–255]. This is a form of restriction applied to the outcome, and in the case of influenza studies, it seeks to mitigate for health seeking behaviour by studying only patients, who have reported symptoms akin to influenza-like illness. A case-control design then analyses effectiveness based on laboratory-confirmed infections, which mitigates against disease misclassification [255]. A poorly defined outcome may evade attempts to control for confounding, in the same way a poorly defined intervention may prove problematic for inferential consistency from the research question and data to a wider population [16]. However, restriction necessarily reduces the external validity of the results. With regards to the test-negative case-control design in study of influenza effectiveness, the size of the study and the generalisability is restricted by access to laboratory-confirmed outcomes [255]. As in the UK and other countries, the absence of widespread laboratory testing to confirm cases of influenza would significantly reduce the sample size. After the initial screening of outcomes, the method still relies on adjustment, through matching on measured confounders, and as a type of case-control study, cannot deal with unmeasured sources.

Another design type is the active comparator new-user study. Because the comparison is between two treatments for the same disease, it can assume the same indicators for treatment with the only difference being the therapy. This ensures a similar distribution of prognostic variables between therapy groups [74]. For that reason, this design could be the optimum design for comparative effectiveness studies, although it is difficult to conceive a wider application of the design beyond such studies. There is no adjustment using longitudinal information per se, but rather eligible patients are selected based on their history of medication use as determined through longitudinal EHRs.

The methodological review of chapter two was scoped to include only studies with at least two exposure levels. In this way, participants would have been recruited according to exposure status. However, this excluded a set of other before-and-after, single-arm, case only designs that recruit according to outcome, in which cases serve as their own controls in a period prior to exposure. One such method is the case series or self-controlled case series, in which cases are selected and their

exposure status retrospectively studied relative to the timing of their outcome [256,257]. Conditional upon the assumption of no effect of a recurrent event altering the probability of subsequent exposure, this approach lends itself well to studying adverse events for drug and vaccine safety. The safety of the mumps, measles and rubella vaccine (MMR) has often been cited as an example of its application [258]. The case series design may be extended to accommodate transient exposures, for which multiple exposure and control periods are possible. This constitutes a case-crossover design [259]. The use of either method depends on the historic pattern of medication use, but both approaches require a well-defined period of exposure and control [215]. For example, after vaccination, there may be a limited window of time during which outcomes could be assumed to be adverse reactions. As the case-crossover design may analyse multiple windows then it is also essential that the baseline rate of such outcomes does not change, and that there are no time trends in the exposure windows. Since individuals in a case only design act as their own controls ("crossover" is within individuals), then this mitigates for fixed confounding effects, but susceptibility to bias from time-varying confounders cannot be ruled out. Furthermore, case-only studies ask a subtly different question of the data, by recruiting cases only. The effect on cases only may be slightly different from the effect adjusted for confounding between exposed individuals and controls, and the results may not be as widely generalisable as those from PERR and Pairwise. Crucially the availability of controls and an exposure-free prior period under the assumption of temporally stable bias allows an investigation of the source and size of bias, in a direct way that is not possible with case only studies.

The methodological review of chapter two broadly considered how longitudinal information could be utilised in the adjustment for unmeasured confounding and was motivated by the growing interest in routinely collected datasets, where such information is likely to be available. Longitudinal information may facilitate the use of "before-and-after" methods, in which longitudinal comparisons are made to temporally adjust for unmeasured confounding. A summary of these methods is listed in Table 4. The inclusion of single-arm, case only methods along with the PERR, Pairwise and other methods reviewed in chapter two further illustrates the overlap between method or study design in adjusting for unmeasured confounding and how design is integral to this adjustment.

Quasi-experimental methods typically invoke extra assumptions to adjust for unmeasured confounding, and before-and-after approaches demand longitudinal information and particular study designs (i.e: cohort) to yield that information. In this way, there is an unavoidable overlap between design and method. Similarly adjustment for unmeasured confounding does not necessarily preclude explicit adjustment for measured confounders (Figure 9). Moreover, since it has been shown that the PERR method cannot entirely remove bias from effect estimates [192], then inclusion of covariates and potential confounders, where available, would be desirable. On this basis, it would be good practice to try and adjust for common confounders, such as age and gender, in any given model first. This would minimise the degree of confounding bias, and leave the PERR method to adjust for residual bias. Furthermore, it would allow assessment of the sources of bias, and of how much is due to unmeasured confounders through comparison of the results from the model adjusted for measured confounders with the PERR-adjusted model. Alternatively, adjustment may take the form of inclusion of a propensity score, rather than adjustment by each variable. Propensity scores were calculated for the studies in this project and more details can be found below in section 4.6.3. Propensity scores may be implemented as an adjustment for observed confounders by more than one approach [212]. For example, inverse probability treatment weights (IPTW) can be derived from propensity scores (PS) to provide estimates from a weighted model. This was the approach followed for explicit adjustment in this project (for more details about the advantage of IPTW over other PS methods, see below in section 4.6.3). Models that re-weight according to measured confounders are also known as marginal structural models [12], although throughout this project, this application has mostly been referred to as IPTW. Weighting may also be applied through a raft of methods broadly known as G-methods to adjust for time-dependent effects if there is information available on past covariates and treatment states [260,261].

| Method | Description | How method controls for unmeasured confounding |
|---|---|---|
| **Prior event rate ratio (PERR)** | Adjustment of the study incidence or hazards rate for rate or survival data, by exposure group, using the equivalent estimate from an exposure-free prior period. | The method relies on individual level data from both periods to control for unmeasured confounding, but adjustment is made at the group level. |
| **PERR-Alt / Pairwise** | Estimates for survival data the hazard rate for an exposure from the paired likelihood of the within-individual comparisons of the study period with an exposure-free prior period. | Adjustment is made within each individual and offers more flexibility in accommodating proportional differences between the period-specific baseline hazards, modelled as a nuisance variable, with the potential to expand the Pairwise formula to more than one period, as well as to model time-dependent covariates. |
| **Difference-in-differences (DiD)** | Estimates the effect on two groups from the coefficient of the interaction between the group and period variables from a regression of outcomes for each group at each time point. | The data look back to an exposure-free period, although DiD does not exclusively require longitudinal data (i.e: may use repeated cross-sectional panel data), and the adjustment is between exposure group. Therefore within-individual data may benefit from including variable specific covariates in the regression, or using the first differences or fixed effects approach. |
| **Fixed effects and first differences** | Estimates the effect on a continuous outcome from regression on a binary exposure having either demeaned the data across time (fixed effects) or by finding the differences (fixed effects method is equivalent to first-differences for two time points). | Requires individual-level data. By regressing the first differences or demeaned data, time-invariant individual-specific confounders are eliminated within each individual. Suitable for continuous outcomes. |
| **Regression discontinuity** | The effect of an intervention may be estimated through the pretest-posttest design, for which a cut-off on the pretest value of a continuous outcome determines membership of a binary intervention group. The effect is estimated as the degree of discontinuity by a dummy variable for treatment in a regression model. For individuals close enough to the cut-off, exchangeability and a lack of confounding bias is assumed. | The method intrinsically requires individual level data to regress post-test on pre-test outcomes. Unmeasured confounding is controlled by assuming that individuals are similar in all potential confounding variables for pretest values within a certain distance from the cut-off. This is best achieved if there is random noise in the pretest measurements, and there is a sharp cut-off in treatment assignment. |

| | | |
|---|---|---|
| **Interrupted time series (ITS)** | Studies the effect of an intervention at a point in time in a population by estimating any discontinuity in the response or change in time-related trends of the response. | ITS may be applied to individual-level longitudinal data or population-level data collected over time, but yields population-level estimates. Accuracy may depend on modelling trend, seasonality or any periodicity within the data, but does not require explicit adjustment for confounders or a control group, although the availability of a control group would allow confirmation of a reported effect. Suitable for studying policy implementations in time, which do not coincide with a concomitant change in any other determinants of outcome. |
| **Self-controlled case series (SCCS)** | Single-arm case only design that recruits on outcomes, and adjusts within individual. | As a single-armed study design, individuals act as their own controls. An observation window should be clearly defined within which it should be possible to identify any possible exposures, which should not be exclusively distributed among cases (i.e: possible to have exposed and unexposed cases). Controls are any events occurring outside a well-defined exposure risk-window. |
| **Case-crossover** | Same as the SCCS, but can accommodate multiple exposure-risk windows for transient exposures. | Same as SCCS, but for multiple exposure-risk windows. |

*Table 4:* A list of before-and-after methods for adjusting for unmeasured confounding with a brief description of each method and how the adjustment is made for unmeasured confounding.

## 4.5 Applications of PERR and Pairwise methods in the literature

Weiner et al. began the first of two studies in 2008 introducing the PERR method by replicating as many aspects as possible of the Scandinavian Simvastatin Survival Study (dubbed the 4S RCT), except for randomisation, using data from the CPRD (then known as the GPRD) [242]. For two of the four outcomes, that the authors were able to replicate, the results of all analyses of MI outcomes showed reasonable concordance with the RCT results. In the case of coronary revascularisation, the PERR-adjusted estimate was not significant and was different from the RCT result, which had demonstrated a beneficial effect. However the PERR-adjusted result was closer to the RCT estimate than the unadjusted results, which indicated a harmful effect. The study published immediately after by Tannen et al., in the following pages of the same issue of the journal replicated two RCTs (HOPE and EUROPA) analysing the effects of angiotensin-converting enzyme inhibitors on five outcomes [243]. Two of the outcomes, death and coronary heart failure could not be PERR adjusted, as such outcomes could not logically be repeated in the period prior to study. Of the three remaining outcomes (myocardial infarctions, strokes and coronary revascularisation) studied in the two trials, the PERR adjustment was demonstrated to be reasonably successful in bringing the estimates into close agreement with the RCT results, although perhaps less for that of stroke under the exclusion criteria used in the HOPE trial.

Later, in 2009, Tannen et al. published the results from the application of PERR to the data from their previously replicated studies [245]. Three of these were the 4S, HOPE and EUROPA replications, which had had the PERR method applied, in previous published work [242,243]. Of the remaining, the PERR method could not be applied to the replication of the Syst-Eur trial [262] studying the effect of antihypertensive therapy because of the method of patient selection, while the PERR adjustment of the replication of the Womens' Health Initiative [263,264] (both those with an intact uterus and the hysterectomy sub-group) was relatively inconclusive because of imprecision and wider confidence intervals. Furthermore, for the point estimates of the effect on strokes, the PERR-adjusted results were further away from those of the RCT and the naïve Cox estimates.

Following up their PERR-adjusted replications of RCTs, Tannen et al. conducted one more study applying the PERR method in comparative effectiveness research of adverse cardiac events under two different thiazolidinedione, antidiabetic drugs: pioglitazone (PIO) and rosiglitazone (ROS) [189]. While the effect of PIO on cardiac outcomes did not differ significantly from published RCT estimates, there was some evidence of deviation from the RCTs in the adjusted HRs and PERR-adjusted HRs for ROS on myocardial infarction (MI). However, there was no statistical evidence of difference in the rate of adverse events between the two drugs, although a direct comparison of the exposed groups revealed a hazard of MI that was higher in the ROS group and borderline significant.

Since the publication of the PERR method and its application in seminal and subsequent papers by its authors, other examples of its application by other authors have only just begun to be published, while the literature review found no examples citing the Pairwise method (Table 5). The first by Brophy et al., captured in the method review of chapter two, examined the effect of proton pump inhibitors (PPIs) on survival times until infections by campylobacter and salmonella [190]. Using the prior event rate, the authors were able to determine that patients, who were prescribed PPIs, were prone to such infections, and not significantly improved by treatment. Since the methodological review of chapter two, Othman et al. have also published a study into PPIs, which through the application of the PERR method showed a reduced risk of pneumonia among the PPI patients [265]. Results were also presented alongside those from an adjusted model, revealing the extent of unmeasured confounding. However, a recently published investigation by Zirk-Sadowski et al. into the effect of PPIs on pneumonia showed an elevated risk of pneumonia in the second year after PPI prescription among adults aged at least 60y using both weighted and PERR-adjusted models [266]. Similar results in the same study were also achieved through the first example outside the seminal study of Yu et al. of the application of the PERR-ALT method.

A novel application of the PERR method was found in Dennis et al.'s study of adverse events following antipsychotic prescribing in older patients with dementia [267]. This was one such example of how the PERR method could be applied to observational, registry data to study a clinical question regarding adverse events in a population that would otherwise be difficult to recruit to trial. Another later study by

Brophy et al. reported higher rates of hospital admissions for respiratory illnesses in children, who were prescribed antipsychotic medication, having applied the PERR method to Welsh EHR data from the SAIL databank [268].

The first example of the application of the PERR method to the estimation of vaccine effectiveness was published in a study by Young-Xu et al. [269]. Using the PERR approach, a beneficial relative effectiveness of high dose vs. standard dose influenza vaccination was reported against influenza and pneumonia-associated outcomes and against all-cause outpatient visits in adults from an US database aged at least 65y, but not against laboratory-confirmed influenza. The point estimate of the latter result would suggest a beneficial effect, but the effect suffered from imprecision and wider confidence intervals. With the availability of laboratory-confirmed results, the data perhaps could have been analysed using a test-negative case-control design, but the PERR method was nevertheless utilised, perhaps justified by the sparsity of laboratory outcomes. However, the authors' use of the period of low circulating influenza viruses immediately before the peak season for the prior period rather than the previous season is questionable. The low-circulation period may be used in the calculation of excess deaths during the peak influenza season [270], and has previously been used in a case-centred model resembling a case control study [271], but in the application of the PERR method, the prior period should be similar in conditions to the study period in order for the analysis to be properly valid as stated above in paragraph 1 of section 4.3. Given the difference in rates and the seasonality of other factors as well as influenza, it is difficult to conceive that the effect of confounders and any interactions with the intervention would be the same between low and high influenza-circulation periods. It is therefore difficult to be assured that the low circulation period would necessarily be representative of the bias during high circulation. If nothing else, then the point estimate from a period with low levels of influenza in circulation would likely be subject to a greater degree of relative imprecision that could lead to an inaccurately adjusted PERR estimate, when a previous influenza season might have be used instead.

Latterly, further work based on the PERR method was presented by Tannen and Yu, called the post-treated event rate ratio (PTERR) [272]. This sought to widen the applicability of the PERR method to studies where mortality is an outcome. As already discussed in section 4.3, being a once-only event, death violates the

condition of repeatability in PERR. The method relies on a post-treatment unexposed duration being available for the exposed group, so may not suit curable conditions requiring single treatment. The authors proposed limiting the as-treated durations of the unexposed group to those of their matched exposed subjects, which could be adjusted using the “post-treated” period when the exposed transition to an unexposed state. While this method with its adjustments for analysis time may at first seem less intuitive than the PERR method, the extensive comparisons of results from database studies to RCTs and simulation results seem supportive of the method as an adjustment for unmeasured confounding in mortality studies. However, the method seems to rely on a low mortality rate and on the bias due to differential case fatality being close to zero or relatively negligible compared to the bias in the as-treated period. These assumptions and the condition of an exposure-free period being available for the exposed group following treatment is likely to limit the applicability of the method. Also, the assumption of temporally stable confounding still applies, and the authors’ emphasis was on the method as a sensitivity analysis and evaluation of unmeasured confounding, rather than as a quasi-experimental method for estimating treatment effects without confounding bias.

| Title | Authors | Year | Journal | Adjustment methods |
|---|---|---|---|---|
| Replication of the Scandinavian Simvastatin Survival Study using a primary care medical record database prompted exploration of a new method to address unmeasured confounding | Weiner MG<br>Xie D<br>Tannen RL<br>Xie D<br>Tannen RL | 2008 | Pharmacoepidemiology and Drug Safety | PERR applied to replicated studies of matched patients |
| Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: further empiric validation of the “prior event rate ratio” to adjust for unmeasured confounding by indication. | Tannen RL<br>Weiner MG<br>Xie D<br>Weiner MG<br>Xie D | 2008 | Pharmacoepidemiology and Drug Safety | PERR applied to replicated studies of matched patients |
| Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: Comparison of database and randomised controlled trial findings. | Tannen RL<br>Weiner MG<br>Xie D<br>Weiner MG<br>Xie D | 2009 | BMJ | PERR applied to replicated studies of matched patients |
| Prior event rate ratio adjustment: numerical studies of a statistical method to address unrecognized confounding in observational studies. | Yu et al. | 2012 | Pharmacoepidemiology and Drug Safety | PERR and PERR-Alt applied to simulated data |
| A new “Comparative Effectiveness” assessment strategy using the THIN database: Comparison of the cardiac complications of pioglitazone and rosiglitazone | Tannen RL<br>Wang X<br>Yu M<br>Weiner MG<br>Wang X<br>Yu M<br>Weiner MG | 2013 | Pharmacoepidemiology and Drug Safety | PERR applied to replicated studies of matched patients compared with propensity-score matched analysis |

| | | | | |
|---|---|---|---|---|
| Incidence of Campylobacter and Salmonella Infections Following First Prescription for PPI: A Cohort Study Using Routine Data | Brophy et al. | 2013 | The American Journal of Gastroenterology | PERR applied to the unadjusted Cox model |
| Community acquired pneumonia incidence before and after proton pump inhibitor prescription: population based study. | Othman F<br>Crooks CJ<br>Card TR<br>Crooks CJ<br>Card TR | 2016 | BMJ | PERR applied to unadjusted Cox models, but compared to adjusted analysis and results from self-controlled case series study design |
| A new method to address unmeasured confounding of mortality in observational studies | Tannen RL<br>Yu M<br>Yu M | 2017 | Learning Health Systems | The PTERR results are compared to results from the unadjusted and adjusted Cox regression as well as those reported from trials |
| Characteristics of Children Prescribed Antipsychotics: Analysis of Routinely Collected Data | Brophy et al. | 2018 | Journal of child and adolescent psychopharmacology | PERR applied to unadjusted Cox models and compared to results from adjusted logistic models |
| Proton-Pump Inhibitors and Long-Term Risk of Community-Acquired Pneumonia in Older Adults | Zirk-Sadowski J et al. | 2018 | Journal of the American Geriatrics Society | The PERR and PERR-Alt method applied to unadjusted Cox models and compared to models adjusted by weighting (inverse probability treatment weighted) |
| Relative Vaccine Effectiveness of High-Dose versus Standard-Dose Influenza Vaccines among Veterans Health Administration Patients. | Young-Xu et al. | 2018 | The Journal of Infectious Diseases | PERR applied to the ratio of incidents calculated before and after matching, and compared to the same results without the PERR adjustment |

Table 5: *List of published studies applying the PERR and Pairwise/PERR-Alt methods*

## 4.6 Settings and implementation of PERR and Pairwise methods in this project

The research question was to investigate the effectiveness of the two vaccines that were the subject of a major health-policy implementation in the UK, of World Health Organisation recommendations at the beginning of the millennium. These were the pneumococcal and influenza vaccine, which are the subject of chapters five and six, respectively. A protocol for these studies, which was submitted to the CPRD, is available in Appendix C – ISAC protocol. Following on from the work of Tannen et al., the intention was to apply the PERR method, and the lesser known Pairwise method, to mitigate for unmeasured confounding in the estimation of vaccine effectiveness. This was compared against an adjusted analysis, using an efficient weighting method in preference to matching, the more commonly encountered method of adjustment used in the replications of RCTs from routinely collected data. The exposure-free prior period would also allow an evaluation of the source and degree of confounding bias.

Comparing the results of the PERR and Pairwise methods is crucial to understanding their relative performance. Furthermore, it is only in the application of such methods that further areas may be identified for future methodological development. In this way, where the interaction of vaccination and age was found to be significant in the study period, the moderating effect of age was also adjusted for confounding bias using the PERR and Pairwise methods. In light of the existing literature, this may be a novel application of the PERR and Pairwise methods.

Pneumococcal and the influenza vaccines have been recommended for adults aged at least 65y, and are intended to ameliorate a major healthcare burden in this age group. However, recent evidence for the effectiveness of the pneumococcal vaccine against coronary disease, and the overlap in the pathophysiology of the diseases that the vaccines are intended to prevent, justifies the multi-faceted approach to the research [273]. There are of course major differences in the pathology of the diseases targeted by vaccination: Streptococcus pneumoniae is a bacterium, against which vaccination is intended to confer long-lived immunity. Influenza is caused by a mix of viruses, which can evolve and change every year and against which the vaccine has

to be matched annually. These differences have ramifications for the study design, and so the vaccine studies require different settings for implementation, which are discussed here. Further information and details in the context of the research, to which they each pertain, can be found in their respective chapters (five and six). However, the common layout of the data from CPRD and similarities in the population, study design and settings justify an initial discussion of these aspects and a description of the workflow that this entails.

### 4.6.1 Study design

In the context of the data, the PERR and Pairwise methods constitute a before-and-after approach to the problem of unmeasured confounding using data, in which longitudinal information is available for the patients selected according to their exposure group, and other prognostic variables. As such, only a cohort study design is applicable. Longitudinal information is arguably available from repeated cross-sectional studies, but in its original form, the PERR method demands that this information is within-patient. Nested case-control studies or test-negative case control studies, as well as case-only studies, also rely on longitudinal information within patients, but here this is from the look-back (purely retrospective) from an outcome as opposed to follow-up (retrospective or prospective) after an exposure.

### 4.6.2 Recruitment and follow-up

Common to both the pneumococcal and influenza vaccine studies, as set out in the method section for each, the minimum age of recruitment was 65y and older at the time of the index date for each study. The date was chosen to coincide with the start of the seasonal rise in vaccination rates, thereby capturing the majority of the vaccinations early on in the study. In this way, most of the follow-up time would take place before the start of the next ‘flu season. For both vaccination studies, recruitment lasted one year. For the influenza vaccine study, follow-up was limited to one year to help retain consistency between the vaccine and the circulating influenza strains. For the pneumococcal vaccine study, however, pathogen evolution was less of an issue, so follow-up was extended to two years, to capture more outcomes whilst still limiting the time during which the effects of confounding might change (see point 1 in section 4.3 on equal net effect of confounding across periods). The follow-up in the prior period was the same as that for the study period, and immediately

preceded the study period. Both studies imposed a vaccine-free period preceding the prior period to allow for wash-out of any previous vaccinations.

In the study of influenza, an estimate of annual vaccine effectiveness was sought over from 1997, which included the first wave of the phased introduction of the vaccine in 1998. Interpretation of the results might therefore accommodate inference about the annual variation in the matching between vaccine and circulating pathogens. In contrast, recruitment to the pneumococcal study utilised the growth in vaccination coverage between 2003 and 2005. Choosing this time period exploited the rapid increase in vaccination rates in a natural experiment to maximise the size of the three annual cohorts (Figure 10). The stepwise introduction of vaccination by age group also facilitated the study effectiveness by these sub groups.

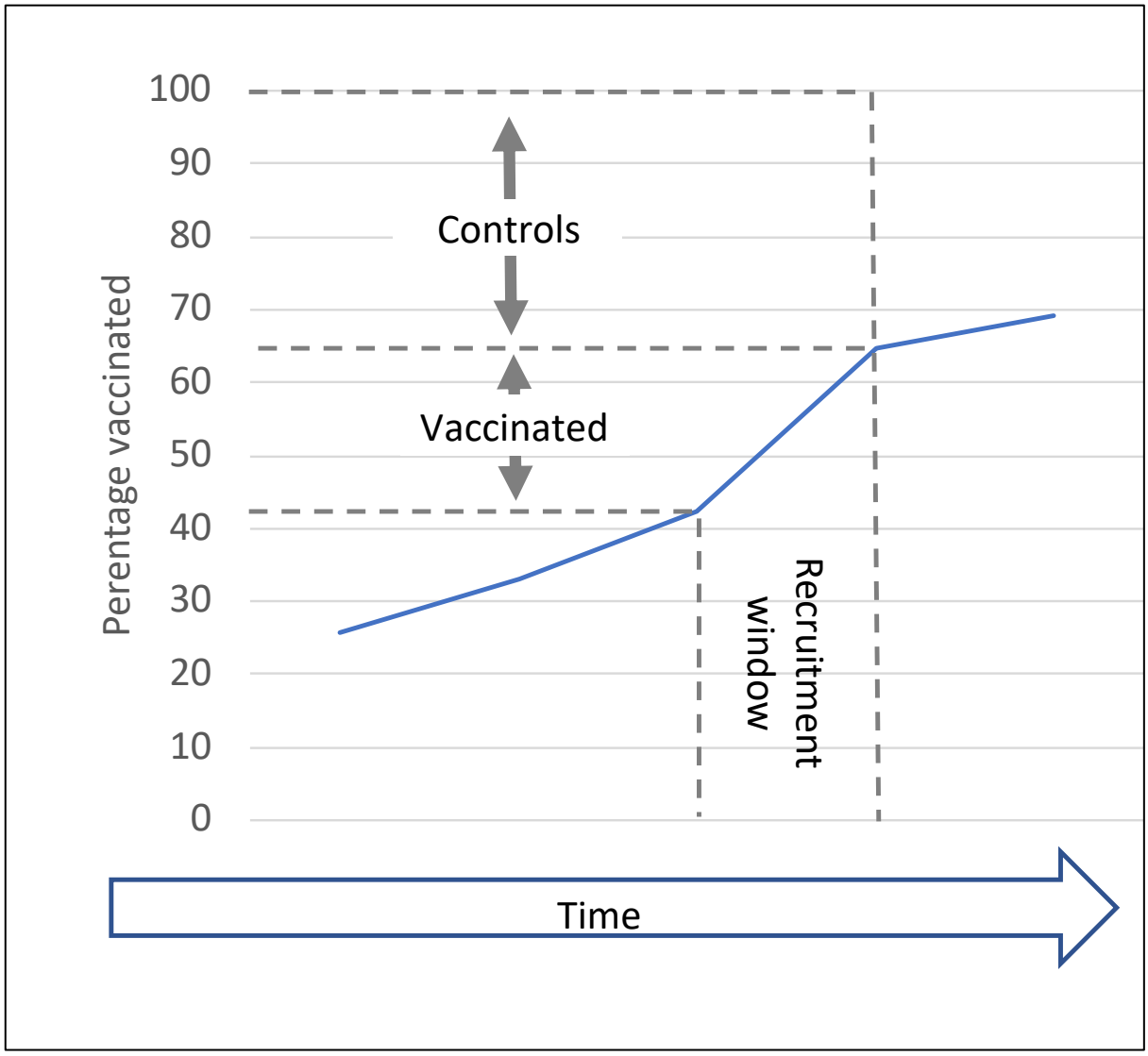


*Figure 10:* Graph of cumulative percentage vaccinated (e.g: pneumococcal vaccination) over time illustrating the advantage for optimal recruitment from growth in vaccine recipients in a population starting with a low vaccination coverage.

The start of the observation times among the vaccine recipients was their date of vaccination. However, no such date exists for the controls. Attributing the cohort index date as a survival start date for all controls would create an imbalance in the distribution of start dates between exposure groups with unforeseen consequences from any time-dependent phenomena or exogenous variables. It is quite possible that some form of immortal time bias could occur among the controls if they all started their observation time much earlier than the vaccine recipients and before any seasonal increase in circulating pathogens. In order to have an approximately equal distribution of observation start times between exposure groups, it was

decided that the start dates for follow-up of the vaccine-free controls should come from the vaccination dates of the vaccine recipients. This was done through matching, at first exactly on age, gender and GP practice, and then iteratively on ever coarser age groups and combinations of gender and GP practice. Where several matches were possible, vaccine dates were randomly allocated as observation start dates to the controls. No further matching was performed, for example, to balance confounders (see below in section 4.7 on the Workflow of computer code).

### 4.6.3 Adjustment for known confounders and effect modification

As a basic adjustment for fundamental patient characteristics, all Cox models, to which the PERR method was subsequently applied, were adjusted for patient age and gender. As such variables are likely to be confounders in their own right, as well as be associated with other potential confounders, this demonstrated good practice, as detailed in section 4.4, in reducing that part of the bias, which the PERR method may fail to adjust. Notwithstanding confounding, the effects of age and gender were of intrinsic interest as prognostic variables. Therefore, these were included in the Pairwise models too.

Since the population under study in this project was older adults, whose life status and disease prognoses are dependent in part on age and age-related frailty, it was of clinical interest to analyse how the effectiveness of vaccination might be moderated by the age of the patients. This entailed fitting an interaction between vaccination and age to the models, and adjusting this using the PERR and Pairwise methods. This was an important and novel aspect of the research, and to my knowledge had not previously been explored using quasi-experimental methods applied to EHRs in this way.

In addition to the quasi-experimental adjustment, it was important to gauge the success, or otherwise, of this through a comparison of a more complete adjustment for the confounders available in the data. To achieve this end, all possible confounders were selected in a logistic regression of vaccinaton status. The predicted probabilities of exposure, or *propensity scores*, provided a summary of the effect of confounders on exposure, reducing potentially many confounders down to a single variable for each individual. However, conditioning on propensity scores in

survival models can lead to biased estimates [274]. Matching introduces minimal bias [275], but relies on the choice of matching algorithm and likely requires more computation to implement. Inverse probability treatment weighting (IPTW) also introduces minimal bias and is recommended for survival models [275], but is easier to implement as this avoids the extra step of matching. IPTW models provide an estimate of the marginal effect of an exposure, which is arguably more readily interpretable than an effect that is conditioned on a variety of prognostic variables. Further details on the prognostic variables selected for the propensity score models, and how these were built, are given in the sections on adjusted modelling within the studies of chapters five and six.

### 4.6.4 PERR estimation

The PERR method can be simply applied to estimate the hazard ratio of vaccine effevctiveness from the ratio of the hazard ratios for vaccination from the Cox model in the study period over that in the prior period. In the two vaccination studies, the Cox regression modelled the time until event on vaccination status adjusted for age and gender, the hazard function is expressed as:

$$h(t) = \exp\left(\beta_x . X + \beta_{age} . age + \beta_{gender} . gender\right) h_0(t)$$

*Equation 2*

where $h_0$(t) is the baseline function and the natural exponent is applied to the linear predictor of the vaccination status indicator variable, X, with coefficients β's for age and gender

The linear predictor in the exponent of the hazard function was specified in the *stcox* command of Stata v.13, and estimated by maximum likelihood. Where interactions between age, gender and vaccination were of interest, these were fitted up to the two-way level and the optimum model selected by comparing the Chi-squared statistics of each Cox model in the study period. In order to apply the PERR method to interactions, the same model specification was applied to the prior period data. Bootstrapping was used to obtain confidence intervals for the PERR-adjusted estimates.

### 4.6.5 Pairwise estimation

Based on the paired Cox model [276] and Yu et al.'s development of the PERR-ALT model [246], Lin and Henley have provided the likelihood expression for the Pairwise model fitted to two periods [192], which was used to fit the model for vaccination status, *x*, adjusting for age and gender:

$$L(X, age, gender, \alpha) = \prod_{i}^{n} \left( \frac{1}{1 + P_i e^{(\beta_x.x + \beta_{age}.age + \beta_{gender}.gender + \alpha)}} \right)^{\Delta_{pi}} \left( \frac{e^{(\beta_x.x + \beta_{age}.age + \beta_{gender}.gender + \alpha)}}{e^{(\beta_x.x + \beta_{age}.age + \beta_{gender}.gender + \alpha)} + S_i} \right)^{\Delta}$$

*Equation 3*

where $\alpha$ is a nuisance parameter modelling the period effect, $\Delta$ is a binary indicator variable denoting whether a survival time for the *i*th of *n* patients in prior period, *p*, or study period, *s*, ended in an event, and *P* is a switch variable denoting whether the shortest survival time was in the prior period, or *S*, the complement of *P* (i.e. *P* = 1- *S*), for the shortest time in the study period.

It can be seen when either a survival time does not end in an event or is not the shortest out of the two periods, then the corresponding term of that period in the likelihood will be unity, and therefore contribute no information to the likelihood. This means in order for a patient to contribute to the likelihood, then the shortest survival time out of the two periods must end in an event, otherwise the information for that patient is null. Effectively this means that the Pairwise estimate is derived from a subset of patients used to find the PERR estimate. This meant less storage and faster computation time when this was deployed in the vaccine studies.

The Pairwise method was applied using R software [277], and the likelihood entered as its logarithm, transforming the product operators into sums. The log likelihood function was estimated using the non-linear minimisation function (*nlm*) in R, with zero starting values and outputting the Hessian matrix. The confidence interval for the Pairwise estimate of vaccination were subsequently derived from the square root of the corresponding diagonal element of the stored Hessian matrix.

### 4.6.6 Cox analysis with time-varying covariates

As an alternative approach to modelling time, a common index date of 1st September of each year was chosen, and vaccination was modelled as a time-varying covariate (TVC). Modelling vaccination as a TVC offered not only insight into any time-dependent effects, but also the advantage of circumventing the need for matching to map the vaccination dates onto the controls for their start times. In this way, chance imbalances in the observation start times between vaccine recipients and controls were avoided, as well as the need to run a computationally intensive matching algorithm in Stata. Cox regression of TVC's first of all required a re-organisation of the data to split the data for the exposed group into two rows of observations, as might be found in the counting process format for the Andersen-Gill model of ordered survival times [278]. For the first row, the exposure indicator variable would be set to zero, to represent the period in an unexposed state, ending in either an event or the time of exposure. The second row, therefore would represent the exposed state of the ordered survival times, with the indicator variable now set to one. The survival time for this second period re-starts from the time of exposure. The controls would remain as they are with one row of observations per individual. If PERR were to be applied, then the format of the prior data would remain as is, with one row of observations per patient. To proceed with TVC Cox analysis in Stata, the data had to be *stset* with the *exit* option specified as "." to retain patient records after failure. The *enter* option was set as the start date for each individual survival time. This was often zero as this corresponded to the common index date, but for post-exposure times, was the time from the start date to exposure (e.g: days since the index date until vaccination). As exposed individuals would contribute two survival times to the likelihood, clustering by patient was accounted for in the covariance matrix of the estimates to account the lack of independence in the exposed observations and to adjust the standard errors accordingly.

### 4.6.7 Sample size calculation

The sample size calculated for both vaccine studies and submitted to the CPRD's Independent Scientific Advisory Committee was partly informed by estimates from the 2012 edition of the Coronary Heart Disease Statistics, from the British Heart Foundation. Focussing on women aged over 84y, this clinically interesting subgroup experienced 139 incidences of acute myocardial infarction (a primary outcome for

ischemic heart disease) per 100 000 person-years. As the meta-analysis into studies of influenza vaccination on cardiovascular disease found the effect could reduce the risk by about 35%, the power to detect a hazard ratio of 0.65 from a Cox model is sought at a power of 0.8 and significance level of 5%. Using the Schoenfeld approximation, this would require a sample size of 121 710 patients. If the power to detect a hazard ratio of, say, 0.70 were sought, this would increase to 177 550 patients.

An estimate of incidence for community-acquired pneumonia among adults aged over 65y put the rate at 7.99 per 1000 person-years from pilot CPRD data. Therefore, the probability of survival times ending in such an outcome over a three-year study period of the pneumococcal vaccine would be approximately 0.024. Results from a meta-analysis of studies into the pneumococcal vaccine suggested that the risk of pneumococcal pneumonia could be reduced by approximately 16%. Therefore, to detect a hazard ratio of 0.85 at a power of 0.8 and significance level of 5% would require 49 530 eligible patients.

## 4.7 Workflow of computer code

The processing of raw data is an essential part of any empirical study, whether a clinical trial or a project using routinely collected data. This involves not only the application of inclusion/exclusion criteria, but manipulation into an analysable form. In this project, as with most studies using CPRD and other routinely collected data, this involved merging together data from across different datafiles, whilst applying the inclusion/exclusion criteria. As both studies in this PhD project investigated vaccine effectiveness and required survival times in a cohort study design, I wrote core code that could be applied to each study with adjustment of files paths, inclusion criteria and the study settings. For clarity, adaptability and ease-of-troubleshooting, this was modularised across a sequence of do-files (files containing program code in Stata), each purposed to apply the inclusion criteria to particular CPRD datafiles, or merge their derivatives and calculate new variables (Figure 11). The modularised do-files were organised in Stata's project manager facility, from which they could be run in order. It should be noted that code writing was an

evolutionary, learning process, so while all code should run without any detectable or undetectable errors, code written later may exhibit more efficiency and arguably less verbosity. While R was used to perform the Pairwise analysis, as it was relatively easy to specify a likelihood within a user-defined function in that particular software, Stata was used for all data preparation and the PERR analysis. This represented a challenge at times as the approach is different from R, plus Stata can only hold one dataset in memory at once, necessitating the specification of temporary datafiles in some instances. Temporary datafiles notwithstanding, many interim datasets were created and saved in between the modular do-files. Graph output was created using both Stata (initially version 12) and the *ggplot2* package in R, depending on convenience.

One particular challenge posed by the constraint of one dataset-at-a-time in Stata's memory was in writing the matching routine stored in the do-file, *Matchexh+dth* (suffixed with the version number). This was used purely for mapping the vaccination dates of the vaccine recipients to the controls in order to assign a similar distribution of observation start dates. No further matching was involved in the analysis, nor as an adjustment for confounding. Thereafter, adjusted analyses were performed through weighting based on derived propensity scores. In the matching routine for observation start dates, the patients were grouped by ever fuzzier levels of grouping variables. First, these were on exact age, gender and practice id. For fuzzier grouping variables, age was categorised into ever coarser levels, until dropped. Practice id was then dropped and further fuzzier levels of grouping organised around just gender and age categorised into varying levels of coarseness, until just gender remained. The do-file routine then looped over the grouping variables, exhaustively matching patients by the finest degree of grouping until moving to the next level of fuzziness. Within the loop, Stata's temporary files were invoked to separate those patients, who had been matched, and keep a tally of the patients remaining. As it is unusual for matching to be used in this way (i.e: just for mapping observation start times), it should be re-iterated that matching was not used in any of the analyes, rather explicit adjustment for confounding was approached through weighted regression.

Stylistic differences may be noticed across the code, with different versions of the same code invoked, as there was an inevitable learning process throughout the

project, which gradually led to slightly more sophisticated use of syntax. In the case of the influenza study, predicting the propensity score for both periods of the 15 cohorts required invoking Mata in the automated selection of the significant prognostic variables for the propensity score models (do file: autoselect_pscore_v2). However, the code writing for this project preceded my recent involvement in NIHR-funded activity focussing on best practice in code writing and programming validation. While this was primarily intended for clinical trials units, there are obvious advantages in terms of traceability and reproducibility in applying the same standards to analysis of EHRs and RCD. Some version control was maintained and Stata's project manager was utilised to run modularised code for this project, however, there are further potential advantages to be gained from using version control software, or uploading to GitHub with its own integrated version-control.

The programming code used in this project accompanies this thesis in separate files.

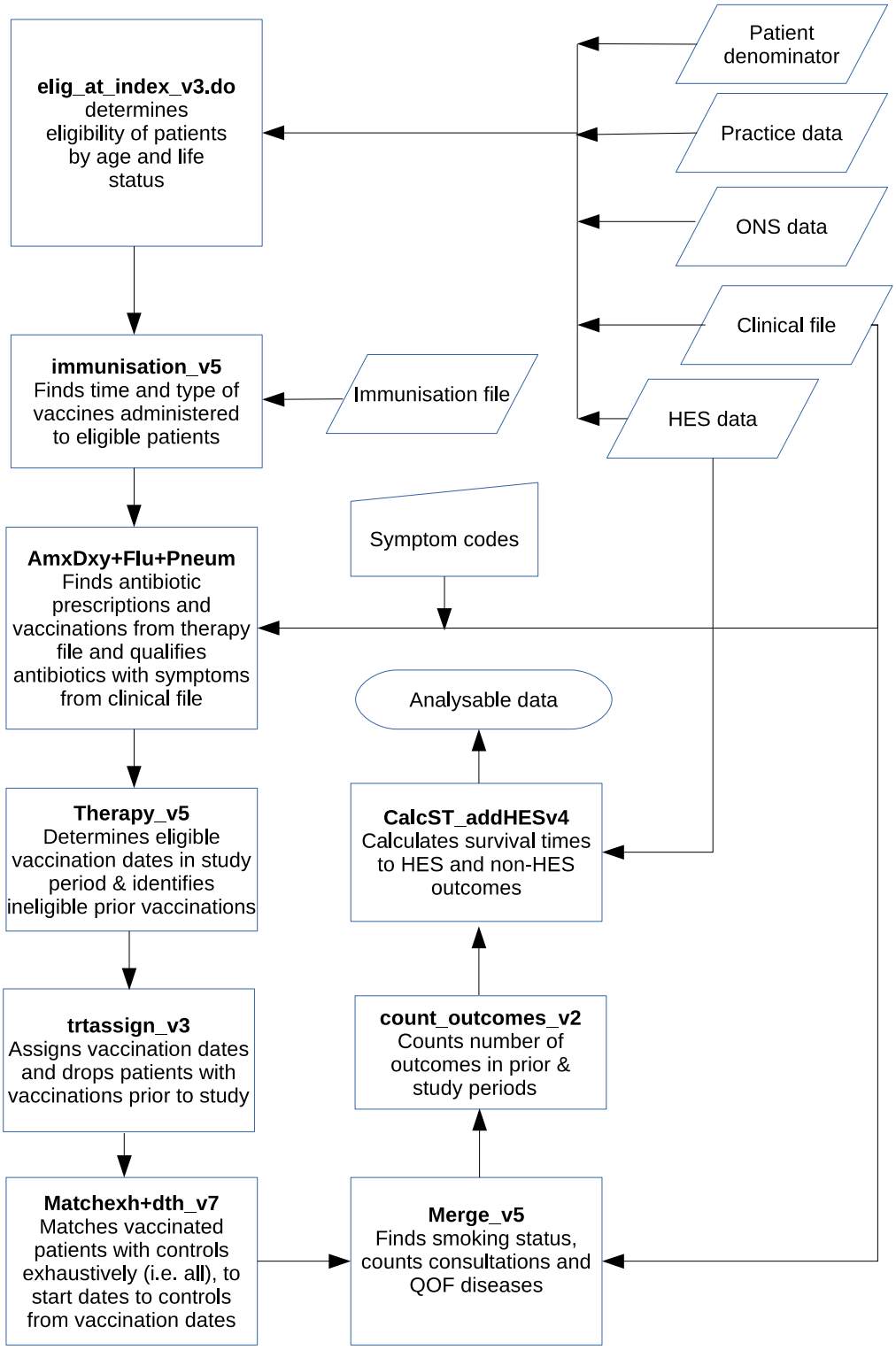


*Figure 11:* Flowchart of data preparation workflow in Stata towards analysable dataset. Do-file names are in bold in square shapes, CPRD datafiles are indicated by rhombus shapes, and datafiles with inputted symptom codes in trapezoid.

# Chapter 5 - Real-world effectiveness of pneumococcal vaccination increases with age in older adults: a quasi-experimental cohort study

Authors:

Adam J. Streeter, Research Fellow in Medical Statistics[a,b]

Lauren Rodgers, Research Fellow in Medical Statistics[a]

Jane Masoli, NIHR Doctoral Fellow and SPR in Geriatric Medicine[c]

Alessandro Ble, Senior Research Fellow in Clinical Epidemiology[c]

Willie Hamilton, Professor of Primary Care Diagnostics[d]

David Melzer, Professor of Epidemiology and Public Health[c]

William E. Henley, Professor of Medical Statistics[a*]

**[a]** Health Statistics Group, University of Exeter Medical School, University of Exeter, South Cloisters, St. Luke's campus, Exeter EX1 2LU, UK

**[b]** Medical Satistics, Plymouth University Peninsula School of Medicine & Dentistry, University of Plymouth, Plymouth Science Park, Derriford, Plymouth PL6 8BX, UK

[c] Epidemiology & Public Health, University of Exeter Medical School, RILD Building, RD&E Hospital Wonford, Barrack Road, Exeter EX2 5DW, UK

[d] DISCOVERY research group, University of Exeter Medical School, University of Exeter, South Cloisters, St. Luke's campus, Exeter EX1 2LU, UK

## 5.1 Abstract


### 5.1.1 Objective

To determine the age-specific effectiveness of the 23-valent pneumococcal polysaccharide vaccine (PPV-23) in UK adults aged 65y and older.

### 5.1.2 Design

Quasi-experimental analysis of a retrospective cohort of 324,804 older primary care patients using the Prior Event Rate Ratio (PERR) method to adjust for measured and unmeasured confounding.

### 5.1.3 Setting

General practices registered to the Clinical Practice Research Datalink with linkage to Hospital Episode Statistics and the Office of National Statistics databases.

### 5.1.4 Population

Three annual cohorts from 2003 to 2005 coinciding with the introduction of the policy to recommend the vaccination against *Streptococcus pneumoniae* for adults aged 65y and older.

### 5.1.5 Intervention

23-valent polysaccharide pneumococcal vaccination (PPV23)

### 5.1.6 Main outcome measure

Specific antibiotic treatment for lower respiratory tract infections or hospitalisation with symptoms consistent with pneumococcal Community Acquired Pneumonia (CAP).

### 5.1.7 Results

For all three cohorts, the rates of pneumococcal pneumonia in the year before vaccination were higher for patients who proceeded to be vaccinated with PPV23 than for patients who remained unvaccinated, indicating the presence of confounding bias. Adjustment for this bias using the PERR method showed that PPV23 was moderately effective for two years after vaccination against CAP in all age sub-groups with hazard ratios of 0.86 (95% confidence interval: 0.80 to 0.93), 0.74 (95%

CI: 0.65 to 0.85) and 0.65 (95% CI: 0.57 to 0.74) in patients aged 65-74, 75-79 and 80+ respectively in the 2005 cohort. The interaction between vaccination and age was statistically significant in the 2005 cohort, with predicted risk reductions of 4%, 12% and 15% at ages 65y, 75y and 80y, respectively.

### 5.1.8 Conclusions

The UK programme for vaccinating patients aged ≥ 65y with PPV23 is effective at reducing CAP. The effectiveness increases with age in step with increasing susceptibility to CAP at older ages. Examining the risk of infection in the period prior to vaccination suggests that the vaccination is targeted towards those most likely to benefit long-term from immunity to pneumococcal infection.

## 5.2 Introduction

Pneumonia is a major cause of morbidity, hospitalization and associated mortality in older adults [279]. Since 2003, public health policy in the UK has recommended vaccination against streptococcus pneumoniae (pneumococcus) for adults aged ≥ 65y using the 23-valent polysaccharide pneumococcal vaccine (PPV23). The vaccination programme began in August 2003 with the PPV23 vaccine offered to adults aged ≥ 80y. This was extended to adults aged ≥ 75y in April 2004 and then finally to all adults aged ≥ 65y in April 2005. PPV23 is recommended as a standard intervention for the elderly in many other countries across Europe and elsewhere. However, there has been ongoing controversy about whether or not PPV23 is effective in preventing noninvasive pneumococcal infection: four systematic reviews have been published since 2016 with divergent conclusions [280–285].

Recent evidence has come from two trials. One, a large-scale, population-based randomised trial (CAPITA) in the Netherlands, reported an efficacy of 46% against first episodes of vaccine-matched strains of community-acquired pneumonia and 75% against invasive pneumococcal disease among 84 496 adults aged ≥ 65y [286]. However, the study lacked the power to draw conclusions on how efficacy might vary with the age of the vaccine recipient. Furthermore, the intervention was protein-conjugated polysaccharides from 13 serotypes (PCV13), a vaccine originally developed for young children but licensed since for use in adults primarily on the basis of immunogenicity studies. PPV23, the vaccine offered to adults in many

countries including the UK, was recently investigated in a prospective, multi-centre trial in Japan, in adults aged ≥ 65y [287]. A significant effectiveness, of 23% and 34%, was reported against all-cause pneumonia and vaccine-matched strains, respectively. However, without randomisation, the trial relied on a test-negative design to mitigate against confounding, with subsequent adjustment in the analysis for confounding variables. The claimed increase in effectiveness in adults aged between 64 and 75y was small, and not supported by statistical evidence. Evidence from previous studies for how the effectiveness of PPV23 changes with age remains inconclusive, yet age is a critical cut-off in determining vaccination policy. Age-related decline in immune function renders older adults susceptible to pneumococcal infection, yet the same decline may reduce the immunogenic response to vaccination.

We conducted a retrospective cohort study using electronic health records (EHRs) to assess real-world effectiveness of the vaccine in adults aged ≥ 65y in the UK and to determine how effectiveness might change with age. The data were extracted from the Clinical Practice Research Datalink (CPRD) with linkage to Hospital Episode Statistics and Office of National Statistics data. Large EHR databases can afford larger sample sizes for the study of real-world effectiveness in small sub-groups than would typically be available in randomised trials, as well as facilitating the study of populations which, for ethical reasons, might otherwise be difficult to recruit into a trial. The rise in vaccination rates resulting from the vaccination programme for older adults provided the opportunity for a natural experiment. Furthermore the incremental introduction of the policy by age group from 2003 to 2005 facilitated estimation of the pneumococcal vaccine's effectiveness within the key age sub-groups.

## 5.3 Methods

### 5.3.1 Data source

We used data from the UK Clinical Practice Research Datalink (CPRD) [53], a database of electronic medical records including information on demographics, consultations, diagnoses, drug prescriptions, immunisations, referrals, etc collected by participating general practitioners (family doctors) during their daily clinical

routines. Our datasets were also linked to hospital admission data and death certificate data. The CPRD has been granted Multiple Research Ethics Committee approval (05/MRE04/87) to undertake purely observational studies, with external data linkages including HES and ONS mortality data. The work of CPRD is also covered by NIGB-ECC approval ECC 5-05 (a) 2012. Our study gained prior approval by the CPRD Independent Scientific Advisory Committee for MHRA database research (ISAC protocol 14-159).

### 5.3.2 Recruitment and study population

Three cohorts were studied, each relating to a single year of the phased introduction by age-group of the policy to vaccinate older adults with PPV23: adults aged ≥ 80y, vaccinated for the first time from 1st September 2003 to 31st August 2004; adults aged ≥ 75y from 1st September 2004 to 31st August 2005; and adults aged ≥ 65y from 1st May 2005 to 30th April 2006. The dates for the start and end of recruitment for each cohort were chosen so as to capture the peak uptake of the vaccine during the relevant period. Patients had to be alive and registered at their general practice at the (index) date of vaccination. All adults that remained unvaccinated for the duration of the study period were designated as controls and matched to vaccinees, to the nearest age, and where possible the same gender and practice, solely for the purposes of assigning an index date. The index date for each control was the vaccination date of the corresponding vaccinee. Each cohort was analysed separately. Any patients without any data in the six years preceding recruitment to the study were excluded from the cohort, as there was a considerable risk they had left the area, but failed to de-register with the practice (see flowchart in section 5.6 Appendix A).

### 5.3.3 Study design and follow-up

Without randomisation, vaccination status in observational studies of this type may be influenced by unmeasured confounders, including variables related to the number and severity of diseases, and in this age group, latent frailty. Recent advances in quasi-experimental methods make it possible, under relevant assumptions, to address directly the unmeasured confounding bias that arises when relevant confounders are omitted [208,288]. One such approach to enhancing the validity of

observational studies is the Prior Event Rate Ratio (PERR) method, proposed by Tannen and Weiner et al [187,242], and extended by Yu et al [246] and Lin and Henley [192]. The PERR approach is becoming more widely adopted in clinical studies based on EHRs and, for example, was used recently to demonstrate that previously reported associations between the use of proton pump inhibitors and risk of community acquired pneumonia are likely to be due entirely to confounding factors [265].

We made use of the PERR framework by considering the introduction of the pneumococcal vaccination policy as a natural experiment. Quasi-experimental analysis of vaccine effectiveness was conducted by using group differences before the introduction of the vaccine to adjust for unmeasured confounders. We adopted a two arm before-and-after design in which vaccinated and control patients were followed up during two periods: the study period, consisting of up to two years from the index date, and a prior period of up to two years starting from two years before the index date (Figure 12). Patients were censored upon death or being transferred out of their practice. The two year study follow-up period was chosen because two years is the time interval for which the effect of the vaccine was found to be stable in previous studies[289] suggesting reasonable stability of unmeasured confounding effects over this period.

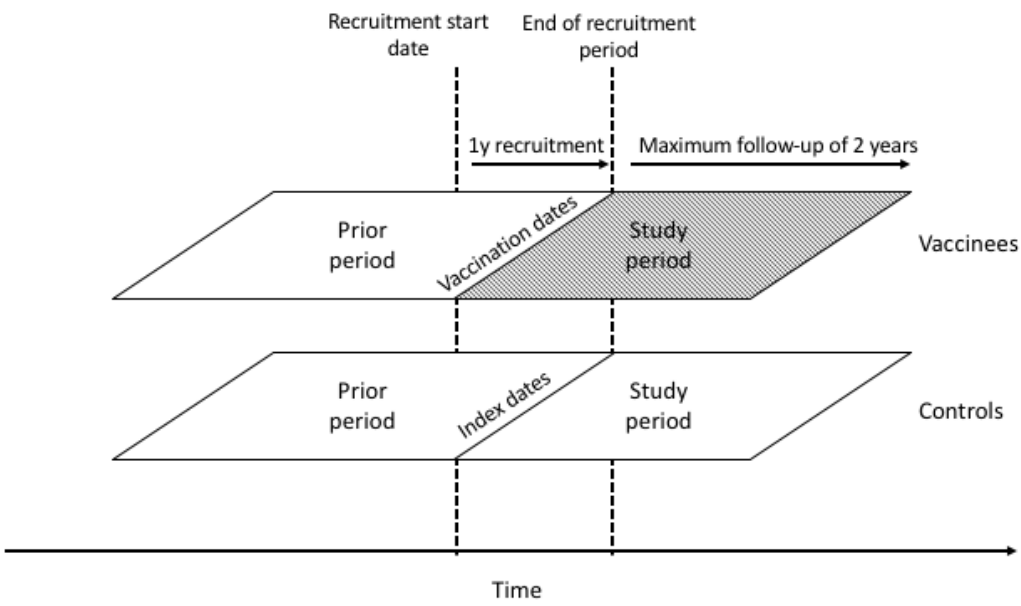


*Figure 12: Schematic of paired design for PERR analysis of PPV23 effectiveness. Patients vaccinated during a 1y recruitment window are selected and matched to controls by age, gender and general practice. Index dates of controls are mapped from the vaccination dates of vaccine recipients. Event times are compared for vaccinated and control patients during a 2y study period and a 2y prior period. The start of the prior period precedes recruitment by exactly 2y. Survival times may end with an event or be censored before the end of either period.*

### 5.3.4 Vaccination

Patients were identified as being vaccinated with PPV23 using relevant codes in the CPRD immunisation file, supplemented with codes from the therapy file (Appendix D – CPRD and HES codes). Similarly, the patients' influenza vaccination status in each year was also retrieved, and included in the study.

### 5.3.5 Outcomes and covariates

As laboratory confirmation of suspected cases of pneumococcal pneumonia is not widely available, we utilized the clinical records to specify a composite outcome of either hospitalisation for suspected pneumococcal pneumonia, or the prescription of antibiotics of species typically used for treating pneumonia, and qualified by the coded symptoms consistent with those of the disease [290,291]. Amoxicillin and doxycycline were the antibiotics selected for study as those most likely to be

prescribed to treat invasive pneumococcal disease and pneumonia-like LRTIs. A pharmacist and clinician independently identified the corresponding product codes (Appendix D – CPRD and HES codes) in the data. The antibiotic prescriptions were further qualified by the medical codes for symptoms consistent with pneumonia and LRTIs, which were independently selected by two clinicians (JM & AB), with input from a third (DM) where differences arose. Hospitalisations were coded according to their ICD-10 classification (Appendix D – CPRD and HES codes). Survival times for this endpoint were calculated for each patient in both the prior and study periods.

Information was also collected on smoking history and comorbidities within the Quality Outcomes Framework (QOF), a scheme to incentivise general practitioners to register common morbidities of patients. The corresponding codes were obtained, with reference to previous work on multimorbidity [60].

For the purpose of building a high-dimensional propensity score predicting treatment, the QOF codes were supplemented with further codes for conditions identified in an electronic frailty index (eFI) [292]. When the codes of a disease category from either QOF or the eFI overlapped with at least 80% of the other, then the source with the fewer codes was excluded to avoid unnecessary multicollinearity in specifying the propensity score model.

### 5.3.6 Statistical analysis

Patient characteristics and comorbidities were used to identify likely sources of measured confounding. Cox's regression was used to model the effect of vaccination on the hazard function for the composite pneumococcal pneumonia outcome, both with and without adjustment for confounding variables (model fitting found in section 5.7 Appendix B: Fitting the Cox model for pneumococcal investigation). For each of the three recruitment cohorts, separate Cox regression models were fitted to the data from the prior and study periods. For the 2003 recruitment cohort, models were fitted to the single age-group of over-79y olds. For 2004, the analysis was stratified by age-group with separate models fitted for over-79y olds and 75-to-79y old adults, while the 2005 cohort was analysed in three separate age-groups: 65 to 74y old adults, 75-to-79y old adults and over 79-y olds. The degree of confounding bias was assessed by the hazard ratio for the treatment effect from the Cox regression models fitted to the treatment-free prior period. Potential modifying effects of age and gender

on treatment effectiveness were considered through inclusion of age, gender and treatment and their interactions up to the 2-way level. The best-fit models for each cohort were selected through chi-square tests of the deviance between nested models. These models were used as the basis for quasi-experimental analysis using the PERR approach. Conventional adjustment for measured covariates, such as the QOF diagnoses, was undertaken as a sensitivity analysis (see below).

### 5.3.7 Adjustment for unmeasured confounding

We used the PERR adjustment to mitigate for hidden confounding, adjusting the hazard ratio (HR) of the study period with that from the prior period. Bootstrap resampling provided the 95% confidence intervals for the PERR estimate. Recent work has shown that the PERR method has a tendency to attenuate treatment estimates in the presence of hidden covariates and / or censoring [225], and so the quasi-experimental analysis was supplemented with the pairwise Cox likelihood approach (pairwise method) which removes this source of bias [192]. This approach is equivalent to the PERR-ALT method proposed by Yu et al [15]. Initially, separate analyses were conducted for each age sub-group (65-74, 75-79, 80+) within each cohort. Further analyses were conducted for each cohort in which the age sub-groups were combined and interactions between age and vaccination status were considered.

### 5.3.8 Sensitivity analyses

Results from methods mitigating for unmeasured confounding were compared with those adjusted by weighting using variables available in the data and identified as having confounding effect [12,293]. This weighted approach as with all adjusted regression models, relies on potential confounders being fully observed and available in the data, with any bias from remaining unobserved confounders assumed to be negligible [12]. The weights were obtained from a propensity score model, but first, the prognostic predictors of survival time were found through backwards elimination of potential confounders in a Cox regression. Eliminating weak prognostic variables that were poorly predictive of survival avoided potential inclusion of instruments of treatment in the propensity score model, which might otherwise inflate the error for the estimated effect [294]. A high-dimensional approach

was taken, using up to 58 candidate variables, including key patient characteristics of age and gender, as well as diseases registered under the UK Quality Outcomes Framework [295], supplemented with variables used in the UK electronic Frailty Index [292] (National Institute of Health Research), and the index itself. Variables in the frailty index were monitored over the five years preceding each patient's study period. The propensity score was obtained as the predicted probability of vaccination from a logistic regression of vaccination status on the candidate variables, identified as potential confounders. Poor predictors of treatment were excluded through backwards elimination of the candidate prognostic variables at the 5% level of significance. Subsequent adjustment of the analytic Cox model with stabilized inverse probability treatment weights (sIPTW) yielded an estimate of the marginal effect and avoided the bias arising from conditioning on the propensity score as has been demonstrated in simulation studies [275].

The propensity score model used to obtain the weights in the study period was re-estimated for the same variables using data from the prior period. The effectiveness in the prior period was then estimated through a weighted regression to ascertain the success of adjustment under the assertion of stable confounding. This was assessed by the proximity of the estimate to unity. However, such an assertion may not have been reasonable, so hazard ratio of effectiveness from the weighted regression of the prior period was re-estimated using weights based on propensity scores modelled on confounders specifically for the prior period data. The PERR analysis was then applied to the weighted-estimate from the study period, alternately adjusted with both versions of the weighted HR from the prior period. i.e.: the HR for the prior period weighted by

- propensity scores specifically obtained for the prior period, but modelled on the same (static) confounders as already determined for the study period
- propensity scores predicted from a model of confounders found to be significant in the prior period, but not necessarily the same confounders as those used for the study-period propensity score (dynamic).

Further sensitivity analysis was also carried out for sub-groups of various influenza vaccination patterns to examine possible confounding between the effectiveness of this and pneumococcal vaccinations on the outcome. To address the impact of

concurrent or recent influenza vaccination on estimates of PPV23 effectiveness, patients were classified into one of four sub-groups based on whether they had been vaccinated against influenza in either the prior or study period alone or in both periods (never-FV, prior-FV, study-FV, always-FV). The potential moderating effect of influenza vaccination on PPV23 effectiveness by age was assessed by modelling an interaction between age and the PPV23 intervention effect in each flu-vaccine subgroup.

## 5.4 Results

Table 1 shows the trends in PPV vaccination coverage by age group, for the periods before and during the introduction of the national vaccination programme. There was good concordance between the PPV uptake achieved by the end of 2005 in the study data (see Table 6) and national vaccination rates reported by Public Health England (PHE - formerly the Health Protection Agency) for uptake by 31st March 2006 (For age groups ≥ 65y; 75-79y; and 80y and older respectively: Extracted data 64.8%, 70.6%, 68.4%; PHE 64.4%, 68.9%, 68.1%). A flowchart showing the numbers at key stages of selection from the data are shown in Figure 14 of section 5.6 Appendix A.

| | Extracted data per year | | Each year by age group | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Year | Number of patients | % Vaccinated | 65 - 74y | | > 75y | | 75 - 79y | | > 80y | |
| | | | patients | %Vac'd | patients | %Vac'd | patients | %Vac'd | patients | %Vac'd |
| 2002 | 470657 | 25.7 | 234482 | 22.4 | 236175 | 28.9 | 94447 | 30.5 | 141728 | 27.9 |
| 2003 | 504948 | 33.2 | 250851 | 24.0 | 254097 | 42.2 | 100334 | 33.7 | 153763 | 47.8 |
| 2004 | 522963 | 42.5 | 260254 | 26.8 | 262709 | 58.1 | 102906 | 55.8 | 159803 | 59.5 |
| 2005 | 541694 | 64.8 | 267984 | 60.2 | 273710 | 69.2 | 107089 | 70.6 | 166621 | 68.4 |
| 2006 | 553627 | 69.3 | 272516 | 65.0 | 281111 | 73.4 | 109473 | 74.6 | 171638 | 72.7 |

*Table 6: Annual cumulative pneumococcal vaccination rates from 2002 to 2006 by age group (for patients alive and registered at the beginning of each year)*

### 5.4.1 Cohort characteristics

The cohort size increased with each study year as the vaccination programme was expanded; about half of each cohort comprised vaccinees with 47.1% in 2003, 41.3% in 2004 and 53.2% in 2005. Nearly half the 2005 cohort were males, decreasing to about a third for the older 2003 cohort (Table 2). The controls were at least two years older on average and tended to have fewer males, and this imbalance appeared to be greatest in the 2004 cohort. The vaccination group was found to have a consistently higher prevalence of diseases registered under the Quality Outcomes Framework compared to the controls. The two leading registered comorbidities were hypertension and coronary heart disease. Hypertension prevalence in the vaccinees ranged from 46% in 2005 to 52% in 2004, and in the controls from 37% in 2005 to 41% in 2004. Similarly, coronary heart disease was more prevalent among the vaccinees with 18.5% compared to 15.6% for the controls from the 2003 cohort falling to 11.3% and 10.5%, respectively, in the 2005 cohort. The 2004 cohort exhibited the greatest imbalance in terms of QOF indices, age and gender. The proportion of identified smokers was similar between treatment groups, increasing slightly with each cohort.

In all three cohorts, the majority of patients had been vaccinated against influenza at least once in both the prior and study periods (always-FV). Compared to those not vaccinated in either period (never-FV), the numbers vaccinated in both periods were 31855 vs. 16500 in 2003, 37771 vs. 26667 in 2004, and 103293 vs. 64016 in 2005. The patients receiving the flu vaccine in the prior period only and in the study period only were relatively few in number ranging from 1911 for those in the 2003 cohort receiving a flu vaccine in the study period only, to 17425 in 2005 for the same flu-vaccine subgroup. The rates of influenza vaccination were much higher in the pneumococcal vaccinees than in the controls: At least 80% of the pneumococcal vaccinees received an influenza vaccination in both periods, while for the controls this figure decreased from 30 to 20% with each cohort (Table 2).

| Cohort | | 2003 | | 2004 | | 2005 | |
|---|---|---|---|---|---|---|---|
| Treatment group | | Vaccinated | Controls | Vaccinated | Controls | Vaccinated | Controls |
| N | | 25870 | 29087 | 30028 | 42625 | 104969 | 92225 |
| % males | | 36.7% | 28.5% | 40.4% | 31.8% | 44.7% | 40.3% |
| Age | Mean | 84.5 | 85.9 | 79.3 | 82.3 | 71.6 | 75.1 |
| | s.d. | 4.0 | 4.8 | 4.3 | 5.8 | 5.4 | 8.1 |
| Disease registered under Quality Outcomes Framework | AF | 14.3% | 11.9% | 11.0% | 10.6% | 6.3% | 6.9% |
| | Asthma | 6.9% | 5.2% | 6.9% | 5.5% | 7.3% | 5.6% |
| | Cancer | 9.7% | 8.2% | 9.7% | 8.8% | 8.4% | 8.0% |
| | CHD | 21.9% | 18.6% | 19.4% | 16.7% | 13.3% | 12.4% |
| | CKD | 6.1% | 3.1% | 19.7% | 12.1% | 14.8% | 11.8% |
| | COPD | 5.4% | 5.3% | 5.9% | 5.2% | 4.4% | 4.2% |
| | Dementia | 6.7% | 9.0% | 4.7% | 7.1% | 1.8% | 3.6% |
| | Depression | 9.8% | 9.0% | 9.4% | 8.7% | 8.5% | 7.9% |
| | DM | 6.2% | 4.2% | 8.8% | 6.4% | 8.4% | 6.9% |
| | Epilepsy | 1.3% | 1.2% | 1.3% | 1.2% | 1.2% | 1.2% |
| | HF | 10.3% | 10.7% | 6.0% | 7.3% | 2.3% | 3.9% |
| | Hypertension | 48.5% | 37.5% | 52.1% | 40.9% | 46.2% | 37.4% |
| | Hypothyroid | 9.0% | 7.9% | 9.2% | 8.2% | 7.9% | 7.2% |
| | Mental Health | 1.0% | 1.6% | 1.2% | 1.6% | 1.1% | 1.5% |
| | Stroke | 14.1% | 14.0% | 10.7% | 11.9% | 5.8% | 7.6% |
| Smoking status | Smoker | 15.4% | 16.4% | 20.9% | 20.6% | 23.7% | 25.3% |
| | Not smoker or n/a | 53.4% | 61.4% | 44.9% | 54.1% | 45.2% | 49.3% |
| | Ex smoker | 31.2% | 22.2% | 34.2% | 25.3% | 31.1% | 25.4% |
| Influenza vaccination in both periods | | 90.2% | 29.3% | 88.1% | 26.6% | 80.1% | 20.8% |

*Table 7: Characteristics of study population for each cohort by pneumococcal vaccination status at cohort entry into study period*

### 5.4.2 CAP and mortality rates in each cohort

The overall risk of the composite CAP outcome decreased from 10.2% in the 2003 cohort to 6.9% in the 2005 cohort, reflecting the younger age distribution for the later cohorts (Table 3). Patients were more likely to experience a CAP event in the study period than in the prior period for both vaccinated and control patients: In 2003, the incidence of CAP in the prior and study periods was 8% and 11% respectively for vaccine patients, and 7% and 10% respectively for control patients. These proportions decreased with later cohorts. The proportion of hospitalisations for pneumococcal pneumonia among patients experiencing a prior or study end point tended to be greater for the controls, as high as 49% for those in the older 2003 cohort, while 31% for the vaccinees. As before, these proportions decreased with later cohorts.

Control patients had higher mortality rates than vaccinated patients with 32% of the controls from the older 2003 cohort being censored on death compared to less than half that figure (15%) among the vaccinees. This imbalance increased with each year of recruitment, a trend that was tempered by the inclusion of younger patients in later cohorts, which saw the overall reduction in mortality fall from 19% to 7% per cohort by 2005. Those hospitalised for pneumonia were at the greatest risk of death, particularly in the older 2003 cohort (68% and 81% following hospital admissions in vaccinees and controls, respectively). Consistent with the high mortality rate following pneumonia hospitalisation, the proportion of outcomes resulting in hospitalisation was lower during the prior period than the study period, as patients needed to be alive after the prior period for subsequent selection to the study. In comparison to deaths, there were far fewer censored survival times due to deregistrations from the general practices.

| Cohort | | 2003 | | 2004 | | 2005 | |
|---|---|---|---|---|---|---|---|
| Exposure group | | Vaccine recipients | Controls | Vaccine recipients | Controls | Vaccine recipients | Controls |
| Deaths | Patients censored on death | 9184 (31.6) | 3985 (15.4) | 9995 (23.4) | 2727 ( 9.1) | 12293 (13.3) | 4296 ( 4.1) |
| Transfers out of practice | Patients censored for transferring out of practice | 3891 (13.4) | 1654 ( 6.4) | 4388 (10.3) | 1316 ( 4.4) | 7719 ( 8.4) | 4227 ( 4.0) |
| Outcomes in study period | Patients with CAP outcomes | 2816 ( 9.7) | 2811 (10.9) | 3579 ( 8.4) | 2734 ( 9.1) | 6016 ( 6.5) | 7625 ( 7.3) |
| | Hospitalised pneumonia (% of CAP) | 1371 (48.7) | 858 (30.5) | 1569 (43.8) | 642 (23.5) | 1901 (31.6) | 1055 (13.8) |
| | Deaths during study among hospitalised pneumonia cases | 1115 (81.3) | 584 (68.1) | 1190 (75.8) | 377 (58.7) | 1328 (69.9) | 528 (50.0) |
| Outcomes in prior period | Patients with CAP outcomes | 2125 ( 7.3) | 2166 ( 8.4) | 2836 ( 6.7) | 2314 ( 7.7) | 5143 ( 5.6) | 7020 ( 6.7) |
| | Hospitalised pneumonia (% of CAP) | 529 (24.9) | 269 (12.4) | 608 (21.4) | 233 (10.1) | 841 (16.4) | 428 ( 6.1) |

*Table 8: Description, N (%), of composite CAP outcomes, death and censoring for each cohort from 2003 to 2005*

### 5.4.3 Prior event rate ratio analysis

For all three cohorts, and in each age sub-group, the rates of CAP in the year before vaccination (prior period) were higher for patients that went on to be vaccinated with PPV23 than for patients who remained unvaccinated, indicating the presence of confounding bias (Table 4). In patients aged 65-74 (2005 cohort), the rate of CAP in the study period was higher for patients that had been vaccinated than for patients that had not: HR=1.28 (95% confidence interval: 1.22 to 1.34). However, the imbalance between vaccinated and control patients was even greater in the prior period before either group had received PPV23: HR=1.37 (95% CI: 1.30 to 1.44). Adjustment for measured and unmeasured confounding bias using the pairwise PERR method gave a significant protective HR of 0.86 (95% CI: 0.80 to 0.93) (Table 4 and Figure 2). Similar protective effects of vaccination were seen in the older age sub-groups within the 2005 cohort with HR of 0.74 (95% CI: 0.65 to 0.85) and 0.65 (95% CI: 0.57 to 0.74) respectively in the 75-79 and 80+ age groups respectively. Results for the 2003 and 2004 cohorts were similar for each age sub-group. Estimates from the standard PERR method were consistent with the pairwise estimates but effect sizes tended to be closer to the null, as expected.

| Cohort year | Age group | Hazard ratios (95% CI) of Treatment term | | | |
|---|---|---|---|---|---|
| | | Prior | Study | PERR | Pairwise |
| 2003 | 80+ | 1.20 (1.13, 1.27) | 1.00 (0.95, 1.06) | 0.84 (0.77, 0.91) | 0.68 (0.63, 0.74) |
| 2004 | 75-79 | 1.23 (1.14, 1.34) | 1.12 (1.03, 1.20) | 0.90 (0.82, 1.00) | 0.82 (0.72, 0.93) |
| 2004 | 80+ | 1.34 (1.23, 1.45) | 1.07 (0.99, 1.15) | 0.80 (0.72, 0.88) | 0.61 (0.54, 0.69) |
| 2005 | 65-74 | 1.37 (1.30, 1.44) | 1.28 (1.22, 1.34) | 0.94 (0.89, 0.99) | 0.86 (0.80, 0.93) |
| 2005 | 75-79 | 1.27 (1.16, 1.39) | 1.08 (0.99, 1.17) | 0.85 (0.76, 0.94) | 0.74 (0.65, 0.85) |
| 2005 | 80+ | 1.31 (1.20, 1.42) | 1.07 (0.99, 1.15) | 0.82 (0.73, 0.91) | 0.65 (0.57, 0.74) |

*Table 9: Hazard ratios, adjusted for age and gender, presented for sub-group analysis of the prior and study periods, and the PERR-adjusted estimates. Sub-groups correspond to the age groups, which were incrementally targeted for pneumococcal vaccination from 2003 to 2005, namely adults aged over 79y; from 75 to 79y; and from 65 to 74y.*

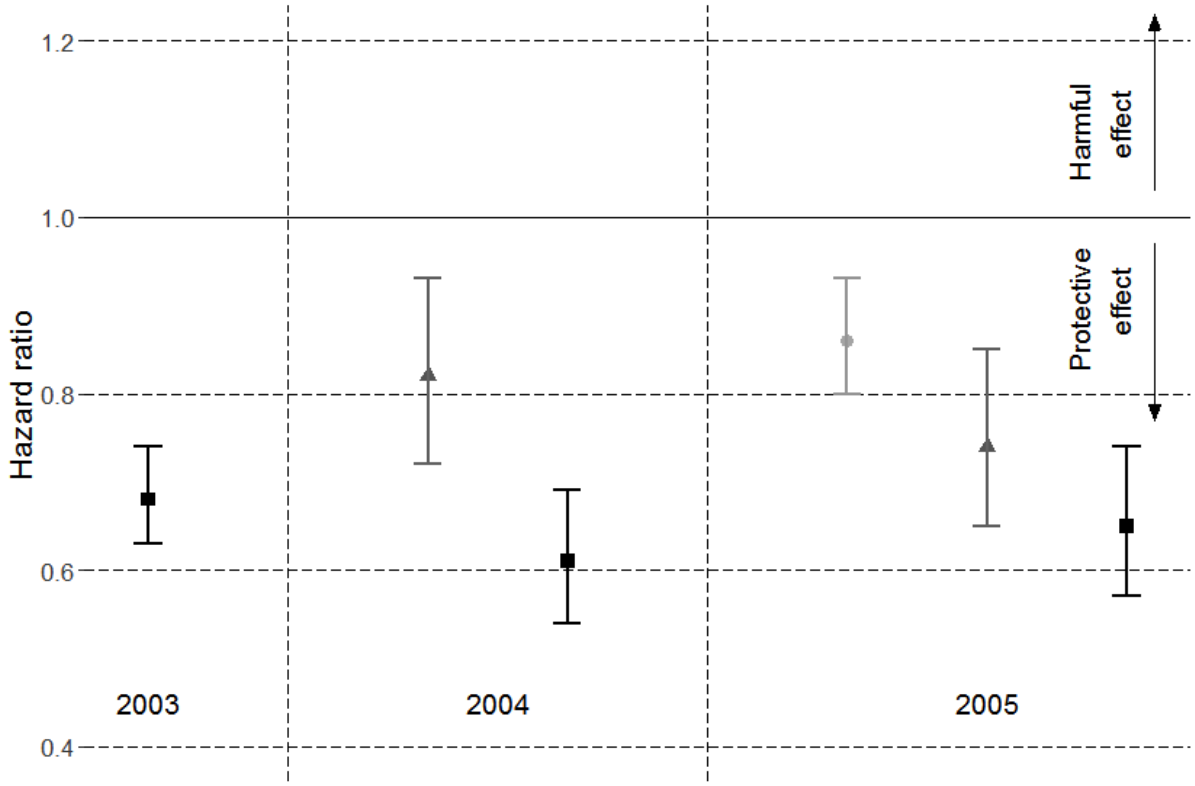


*Figure 13: Pairwise-adjusted hazard ratios of vaccination for each annual cohort (2003-005) by sub-groups of age (65 to 74y – light grey circle; 75 to 79y – mid-grey triangles; 80+y – black squares).*

Further analysis of the 2005 cohort, modelling age as a covariate in an interaction with vaccination, identified an increasing protective trend with age (p-value=0.01). The interaction HRs for the PERR and pairwise models, respectively, were 0.99 (95% CI: 0.99 to 1.00) and 0.98 (95% CI: 0.98 to 0.99) indicating that the estimated reduction in the rate of CAP in the vaccinated patients from each model improved by 4% and 8%, respectively, for every 5 year increase in age.

One of the central assumptions of the PERR adjustment approach is that the occurrence of prior events does not influence the likelihood of future treatment. We note that in this study, outcomes in the prior period did not greatly differentiate subsequent vaccination status, with 50.5% of patients with CAP being vaccinated, compared to 46.8% of those without CAP in the same period.

## 5.4.4 Sensitivity analysis

### *5.4.4.1 Propensity score weighted regression*

The list of confounders found to be significant in the study period of each cohort is given in Table 10, as well as those (dynamic), which were found specifically for the prior period. The balancing variables found not to be significant, and therefore excluded as potential confounders, in models in any period for any of the cohorts were: cancer, mental health, stroke, aspirin count. In addition to these, there were inconsistent confounders, which appeared no more than once in the study period of any of three cohorts: QOF-registered atrial fibrillation, QOF-registered chronic kidney disease, epilepsy, hypothyroidism. More confounders were found in the study periods. In 2003, the disparity in the number of confounders was 47 in the study period compared to 44 in the prior of the 2003 cohort, but this grew to 49 compared to 30 for the same respective periods in the 2005 cohort.

Density plots of the propensity scores (section 5.8 Appendix C: propensity score density graphs) revealed sufficient overlap for valid inference for the 2004 (Figure 22) and 2005 cohorts (Figure 23), but a slight disparity in overlap towards the higher scores and a concern about the positivity of controls, with many scores close to zero, raised caution about inference from the IPTW analysis of the 2003 cohort (Figure 21). The success of the IPTWs to balance the confounders was judged by comparison of the unweighted and weighted standardised mean differences (SMD) of the variables used to predict the propensity score (Plotted as Figure 24, Figure 25 and Figure 26 in section 5.9 of Appendix D: plots of standardised mean differences of balancing variables, for the 2003, 2004 and 2005 cohorts, respectively). Notably, age, then the number of historic consultations and prescriptions, were consistently the variables with larger imbalances, indicating these as the primary confounder variables. For the 2003 and 2004 cohorts, the electronic frailty index was also one the largest sources of confounding. Weighting achieved good balance for the 2005 cohort, with no weighted SMDs exceeding six units (pooled weighted standard deviations), and the largest imbalance due to age being substantially reduced. Reasonable balance was achieved for all confounders in the 2004 cohort, except age, for which the SMD remained relatively high at 30 units, although this was half the unweighted SMD. For the 2003 cohort, the weighed SMD for age was also half

that of the unweighted SMD, but remained high at 70 units. Although the unweighted SMDs indicated the largest imbalances in the 2003 cohort, following weighting the SMDs remained high, though still reduced for many variables.

When analysing the study period data through an IPTW regression (using weights based on the propensity scores predicted from the logistic regression of confounders), the greatest effect of vaccination occurred in the 2003 cohort among the 80+y age group with a HR of 0.72 (95% CI: 0.67 to 0.78). By 2004, this had increased to 0.84 (95% CI: 0.77 to 0.92) for the same age group, but by comparison to the 75-79y age group (HR 0.93; 95% CI: 0.85 to 1.17), there was an apparent age-related effect. By 2005, this direction of effect in the study period has disappeared, with the 75-79y group have the lowest HR at 0.90 (95% CI: 0.83 to 0.98). (

| Cohort | Age group | N vaccinated | N controls | HR for each period | | PERR HR |
|---|---|---|---|---|---|---|
| | | | | Prior | Study | |
| 2003 | 80+ | 25870 | 29087 | 0.89 (0.83, 0.96) | 0.72 (0.67, 0.78) | 0.81 (0.74, 0.89) |
| 2004 | 75-79 | 19409 | 16632 | 1.07 (0.98, 1.17) | 0.93 (0.85, 1.01) | 0.86 (0.78, 0.96) |
| 2004 | 80+ | 10619 | 25993 | 1.09 (0.99, 1.19) | 0.84 (0.77, 0.92) | 0.77 (0.68, 0.87) |
| 2005 | 65-74 | 79812 | 49879 | 1.16 (1.10, 1.22) | 1.03 (0.98, 1.09) | 0.89 (0.84, 0.95) |
| 2005 | 75-79 | 15784 | 16403 | 1.07 (0.98, 1.17) | 0.90 (0.83, 0.98) | 0.85 (0.76, 0.94) |
| 2005 | 80+ | 9373 | 25943 | 1.22 (1.11, 1.33) | 0.96 (0.86, 1.08) | 0.79 (0.69, 0.91) |

Table 11).

The weighted HRs of the treatment groups in the two-year prior period were used to evaluate the degree of unmeasured confounding that might still remain after the weighted adjustment. Comparing the estimates using weights based on (dynamic) confounders specific to the prior (Table 11) and those based on the same (static) confounders found in the study period (

| Cohort | Age group | N vaccinated | N controls | HR for each period | | PERR HR |
|---|---|---|---|---|---|---|
| | | | | Prior | Study | |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2003 | 80+ | 25870 | 29087 | 0.87 (0.79, 0.96) | 0.72 (0.67, 0.78) | 0.83 (0.75, 0.94) |
| 2004 | 75-79 | 19409 | 16632 | 1.08 (0.99, 1.18) | 0.93 (0.85, 1.01) | 0.86 (0.78, 0.96) |
| 2004 | 80+ | 10619 | 25993 | 1.10 (1.00, 1.21) | 0.84 (0.77, 0.92) | 0.76 (0.68, 0.86) |
| 2005 | 65-74 | 79812 | 49879 | 1.17 (1.11, 1.23) | 1.03 (0.98, 1.09) | 0.89 (0.83, 0.94) |
| 2005 | 75-79 | 15784 | 16403 | 1.09 (1.00, 1.19) | 0.90 (0.83, 0.98) | 0.83 (0.74, 0.93) |
| 2005 | 80+ | 9373 | 25943 | 1.26 (1.15, 1.38) | 0.96 (0.86, 1.08) | 0.77 (0.66, 0.87) |

Table 12), the results were similar across the two different approaches to weighting for the prior period, indicatingreasonable stability in the modelling of confounders. The deviation from the null of the prior period HRs indicated that some residual confounding could be assumed and were greater than unity except for 2003, which indicated a bias towards healthy vaccine recipients. Once the PERR-adjustment had been applied to the weighted study-period estimates, an age effect was consistently observed across the cohorts: The HRs for the 80+y group ranged from 0.77 to 0.81 based on dynamic confounders, and from 0.76 to 083 for the static confounders. The point estimate for the 75-79 y age group varied between 0.85 and 0.86 for both the 2004 and 2005 cohorts modelled on dynamic confounders, and between 0.83 and 0.86 modelled on static confounders. For the youngest age group, 65-74y, in 2005, the point estimates were both 0.89 (to the nearest 2 d.p.) regardless of the type of confounders used.

| Disease | 2003 | | 2004 | | 2005 | |
|---|---|---|---|---|---|---|
| | Prior | Study | Prior | Study | Prior | Study |
| Asthma | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Atrial fibrillation | | | ✓ | | | |
| Cancer | | | | | | |
| Coronary heart disease | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Chronic kidney disease | | | | | | ✓ |
| Chronic obstructive pulmonary disorder | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Dementia | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Depression | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Diabetes | | ✓ | | | | ✓ |
| Epilepsy | | | ✓ | ✓ | | |
| Heart failure | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Hypertension | | | | ✓ | | ✓ |
| Mental health | | | | | | |
| Stroke | | | | | | |
| Hypothyroidism | | | ✓ | ✓ | | |
| Activity limited | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Anemia | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Arthritis | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Atrial fibrillation | ✓ | ✓ | | ✓ | | ✓ |
| Carer required | ✓ | ✓ | | ✓ | | ✓ |
| Cerebrovascular disease | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Chronic kidney disease | | | | ✓ | | ✓ |
| Diabetes | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Dizziness | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Dyspnoea | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Falls | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Foot problems | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Fracture | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Hearing impaired | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

| | | | | | | |
|---|---|---|---|---|---|---|
| Heart failure | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Heart valve disease | | ✓ | | ✓ | | ✓ |
| Housebound | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Hypertension | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Ischaemic heart disease | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Mental / cognitive problems | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Mobility | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Osteoporosis | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Peptic ulcer | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Peripheral vascular disease | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Polypharmacy | | | | | | |
| Respiratory disease | ✓ | ✓ | | ✓ | ✓ | |
| Skin ulcer | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Sleep disturbed | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Social vulnerability | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Syncope | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Thyroid disease | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Tremors | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Urinary incontinence | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Urinary system disease | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Visual impairment | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Weightloss and anorexia | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Yearly drug count | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Electronic frailty index | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Age | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Gender | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Aspirin count | | | | | | |
| Number of consultations | | ✓ | ✓ | ✓ | ✓ | ✓ |
| Smoking status | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

*Table 10: List of balancing variables used to predict propensity scores for each cohort 2003-2005. The number of consultations and aspirin prescriptions were counted, respectively, over three and two years preceding the start of each patients follow-up in the study period*

| Cohort | Age group | N vaccinated | N controls | HR for each period | | PERR HR |
|---|---|---|---|---|---|---|
| | | | | Prior | Study | |
| 2003 | 80+ | 25870 | 29087 | 0.89 (0.83, 0.96) | 0.72 (0.67, 0.78) | 0.81 (0.74, 0.89) |
| 2004 | 75-79 | 19409 | 16632 | 1.07 (0.98, 1.17) | 0.93 (0.85, 1.01) | 0.86 (0.78, 0.96) |
| 2004 | 80+ | 10619 | 25993 | 1.09 (0.99, 1.19) | 0.84 (0.77, 0.92) | 0.77 (0.68, 0.87) |
| 2005 | 65-74 | 79812 | 49879 | 1.16 (1.10, 1.22) | 1.03 (0.98, 1.09) | 0.89 (0.84, 0.95) |
| 2005 | 75-79 | 15784 | 16403 | 1.07 (0.98, 1.17) | 0.90 (0.83, 0.98) | 0.85 (0.76, 0.94) |
| 2005 | 80+ | 9373 | 25943 | 1.22 (1.11, 1.33) | 0.96 (0.86, 1.08) | 0.79 (0.69, 0.91) |

*Table 11:* PERR-adjusted analysis of pneumococcal vaccine effectiveness in the 2003-2005 cohorts, based on the inverse probability treatment weighted hazard ratios, with the weights estimated for each period from propensity scores predicted from period-specific (dynamic) confounders.

| Cohort | Age group | N vaccinated | N controls | HR for each period | | PERR HR |
|---|---|---|---|---|---|---|
| | | | | Prior | Study | |
| 2003 | 80+ | 25870 | 29087 | 0.87 (0.79, 0.96) | 0.72 (0.67, 0.78) | 0.83 (0.75, 0.94) |
| 2004 | 75-79 | 19409 | 16632 | 1.08 (0.99, 1.18) | 0.93 (0.85, 1.01) | 0.86 (0.78, 0.96) |
| 2004 | 80+ | 10619 | 25993 | 1.10 (1.00, 1.21) | 0.84 (0.77, 0.92) | 0.76 (0.68, 0.86) |
| 2005 | 65-74 | 79812 | 49879 | 1.17 (1.11, 1.23) | 1.03 (0.98, 1.09) | 0.89 (0.83, 0.94) |
| 2005 | 75-79 | 15784 | 16403 | 1.09 (1.00, 1.19) | 0.90 (0.83, 0.98) | 0.83 (0.74, 0.93) |
| 2005 | 80+ | 9373 | 25943 | 1.26 (1.15, 1.38) | 0.96 (0.86, 1.08) | 0.77 (0.66, 0.87) |

*Table 12:* PERR-adjusted analysis of pneumococcal vaccine effectiveness in the 2003-2005 cohorts, based on the inverse probability treatment weighted hazard ratios, with the weights estimated for each period from propensity scores predicted for each period, but based on those (static) confounders found to be significant in the study periods only.

### *5.4.4.2 Impact of current or recent influenza vaccine*

Separate analysis of the 2005 cohort by influenza-vaccine sub-group indicated that PPV effectiveness was maintained at age 80 irrespective of whether patients received the influenza vaccine in either one or both of the prior and study periods (HR for the pairwise models: 0.69, 0.74, 0.61 and 0.82 for the never-FV, prior-FV, study-FV and always-FV sub-groups respectively in Table 13). The increasing protective effect of the PPV with age seen for the overall 2005 cohort was maintained, irrespective of the confounding-adjustment method, in all flu-vaccine subgroups with very similar gradients (HR for the interaction in the pairwise models ranged from 0.97 to 0.98 in Table 13). We noted that there were inconsistent results for the main effect of vaccination across the 2005 influenza-vaccine sub-groups: patients only receiving the influenza vaccine in the study period (post-FV) was the only sub-group with a protective effect of PPV23 at age 65. The main effects for the other three sub-groups were all above one. Analysis of the 2003 and 2004 cohorts, provided further support for the effectiveness of the PPV23 vaccine at the oldest ages across all four influenza vaccine sub-groups, although the precision of the

estimates was affected by the small number of PPV23 recipients in the never-FV and prior-FV sub-groups, and the small number of controls in the post-FV.

| FV sub group | years | N | | Hazard ratios | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | vaccine | controls | Vac | lcl | ucl | Vac*age | lcl | ucl | Vac @65y | lcl | ucl | Vac @ 75y | lcl | ucl | Vac @ 80y | lcl | ucl |
| Never FV | 2003 | **859** | 15641 | 0.73 | 0.49 | 1.09 | | | | | | | | | | | | |
| | 2004 | 1484 | 25183 | 0.74 | 0.54 | 1.01 | | | | | | | | | | | | |
| | 2005 | 5386 | 58630 | 1.08 | 0.81 | 1.44 | 0.97 | 0.95 | 1.00 | 1.08 | 0.81 | 1.44 | 0.80 | 0.66 | 0.97 | 0.69 | 0.54 | 0.88 |
| Prior FV | 2003 | **527** | 4164 | 0.48 | 0.30 | 0.76 | | | | | | | | | | | | |
| | 2004 | **667** | 5023 | 0.64 | 0.41 | 0.99 | | | | | | | | | | | | |
| | 2005 | 2760 | 9880 | 1.26 | 0.82 | 1.91 | 0.97 | 0.93 | 1.00 | 1.26 | 0.82 | 1.91 | 0.88 | 0.68 | 1.13 | 0.74 | 0.53 | 1.02 |
| Post FV | 2003 | 1141 | **770** | 0.57 | 0.35 | 0.93 | | | | | | | | | | | | |
| | 2004 | 1425 | 1100 | 0.44 | 0.28 | 0.68 | | | | | | | | | | | | |
| | 2005 | 12698 | 4547 | 0.70 | 0.51 | 0.96 | 0.99 | 0.96 | 1.02 | 0.70 | 0.51 | 0.96 | 0.64 | 0.51 | 0.80 | 0.61 | 0.45 | 0.82 |
| Always FV | 2003 | 23343 | 8512 | 0.69 | 0.60 | 0.79 | | | | | | | | | | | | |
| | 2004 | 26452 | 11319 | 0.76 | 0.67 | 0.86 | | | | | | | | | | | | |
| | 2005 | 84125 | 19168 | 1.17 | 1.00 | 1.35 | 0.98 | 0.97 | 0.99 | 1.17 | 1.00 | 1.35 | 0.92 | 0.84 | 1.01 | 0.82 | 0.74 | 0.92 |

*Table 13:* Results from pairwise regression of survival times adjusted for age and gender, modelling main effects of vaccination (Vac) for the 2003-2004 cohorts, and their interaction (Vac*age for the 2005 cohort, by sub groups according to flu vaccination in the prior and study periods. For the interactions of the 2005 cohorts, the predicted hazard ratios at ages 65, 75 and 80y are presented, along with the bootstrapped lower (lcl) and upper (ucl) 95% confidence intervals for all hazard ratios. Number (N) of patients less than 1000 are highlighted in red to draw attention to the small numbers in some cases

## 5.5 Discussion

### 5.5.1 Statement of *principal* findings

The results from this study have shown that vaccination with PPV23 is effective in protecting older adults aged 65 and above against pneumococcal community-acquired pneumonia in routine clinical practice. The conclusion was based on concordance between results, having taken a robust approach to confounding bias, applying different methods to adjust for confounding. To the best of our knowledge, this is the first population study to establish that vaccine effectiveness is maintained, and may even increase, in the oldest age groups: the reduction in risk due to PPV23 vaccination was estimated to be about 15% in adults aged 65-74 and increased to 35-40% in adults aged 80 or above. Aggregating the number of events and the time at risk over the three consecutive years for the group aged 80 or above, the average percentage risk over 365 years in the control group was calculated to be 7.1%. Treating this as the baseline risk, and assuming a relative risk of 0.65 among the vaccinated as estimated for 2005 cohort, this corresponded to a reduction in absolute risk of 2.5%. Hence, the vaccine may be expected to prevent 25 cases of PPV every year among every 1000 adults aged at least 80y.

### 5.5.2 Strengths and weaknesses of the study

Our study has several strengths: firstly, our data source, the Clinical Practice Research Datalink, with current coverage of about 11 million patients, is representative of the general population of patients in the UK [296]. Using this database and adequate sample selection strengthens generalizability of our findings from 324,804 elderly patients. We believe that risks estimated in this study represent actual real-world events during the study periods. Our selection criteria allowed for inference on the general population of older patients aged 65 years and over rather than only those who were at risk of pneumonia. The study therefore has high external validity. The introduction in August 2003 of the UK policy of vaccinating older adults against *Streptococcus pneumoniae* created a natural experiment. By sampling patients during the early years of this programme when uptake of PPV23

was high, we were able to select sufficiently large numbers of patients receiving PPV23 in order to study vaccine effectiveness by age sub-group.

We believe this study is the first example of using Prior Event Rate Ratio methodology to control for unmeasured confounding in a vaccine effectiveness study. We have used a recent formulation of the PERR approach, the pairwise method (also equivalent to the PERR-ALT method) that overcomes a source of bias in the original PERR adjustment method, by fitting a paired Cox model to the prior and study periods. It has been demonstrated that applying PERR methods to retrospective cohorts, under certain conditions, can reproduce results from randomised clinical trials [187–189,242]. However, PERR methodology is limited by the need for stronger and more complex assumptions than randomisation. Firstly, vaccination should not be determined by presentation of the outcome in the prior period. This assumption is likely to have been met given that we found little difference in vaccination status between patients with a suspected pneumonia event in the prior period and those without. The second main assumption is the lack of substantive time-dependent confounding. We tried to address this by limiting the follow-up to two years post-vaccination and by replicating results for the 74-79 and 80+ age sub-groups across multiple recruitment cohorts. We also made comparisons with estimates from standard Cox regression models weighted by high-dimensional propensity scores, a well-established approach to dealing with measured confounding in observational studies. The gold standard for evidence remains a well powered RCT, as trials are able to remove both the influence of time-invariant confounders (e.g. associated with genetic variance) and time-variant confounding (e.g. associated with a temporary health state) from the analyses. However, RCTs are not always representative of the clinical populations of interest, especially the oldest old and frail patients that may be the target for a health intervention.

Another concern was over the potential repeatibility of the outcomes. While repeated pneumonia infections may be possible, particularly among frail, older adults in this population, infection may provoke an immune response, that may confer some immunity against later infection, and so change the risk. If this were the case, then one could expect to see a reduced hazard of further infection relative to the prior period, particularly in the unvaccinated patients of the control group. This would represent a violation of the assumption of period-invariant confounding. In the case

such as this where there is a greater, pre-exisiting risk among the vaccine recipients, then the PERR-adjusted effects would be exaggerated.

An important limitation of our study was the lack of information on pneumococcal pneumonia serotypes. The choice of a composite outcome measure based on antibiotic prescriptions or first hospitalization for suspected pneumococcal pneumonia was less specific than in some previous studies but was developed with clinician input to reflect the manifestations of pneumococcal disease in clinical practice.

While our study addresses real-world effectiveness of pneumococcal vaccination up to two years post-vaccination, questions remain over the long-term immunity afforded by PPV23. Although stability of confounding factors may be reasonable over the short term, a longer follow-up of more than two years would inevitably capture declining health and increases in frailty. Where changes in the confounding relationships are time-dependent, the assumptions of many quasi-experimental analysis methods, including the PERR approach, would be violated.

### 5.5.3 Strengths and weaknesses in relation to other studies

Until recently, much of the evidence for the efficacy of the PPV23 vaccine has been based on studies in younger, healthier adults. Of two reviews in 2016/2017 focusing on older adults, the review by Schiffner and colleagues concluded that there was no evidence that PPV23 can prevent CAP in a general, community-dwelling elderly population[284] . In contrast, the review by Falkenhorst and colleagues reported significant vaccine efficacy/effectiveness against both IPD and pneumococcal pneumonia[285]. The two reviews identified the same RCTs and the difference in findings relates to decisions over inclusion criteria and the quality of evidence provided by each study. This lack of consistency between systematic reviews has led to ongoing controversy surrounding the effectiveness of the PPV23 vaccine.

Although the recent CAPITA trial has shown the efficacy of the 13-valent pneumococcal conjugate vaccine against pneumococcal pneumonia and invasive pneumococcal disease in adults aged 65 years or older, it did not resolve uncertainties surrounding effectiveness of PPV23 [286]. The recent multicentre, prospective study conducted in Japan by Suzuki and colleagues found low to

moderate effectiveness of PPV23 against vaccine serotype pneumococcal pneumonia[287]. Compared with conventional case-control or cohort designs, their test-negative study design was less susceptible to bias caused by differences in health-care-seeking behaviour among cases and controls, and the use of non-specific outcome measurements. However, it is not clear how well this approach controls for general sources of unmeasured confounding. The generalisability of the results was also restricted to those patients presenting with symptoms of CAP, for whom laboratory confirmation was available, and the study lacked the power to look at the important question of how PPV23 effectiveness varies by age group. In contrast, our study employed two methods to adjust for unmeasured confounding, and compared these with a high-dimensional adjustment for measured confounders, across subgroups of age. Ours is the largest study to date and made it possible to compare the effectiveness of PPV23 across age sub-groups. The finding that vaccine effectiveness may increase with age reflects the increased vulnerability to infection of the oldest old. In contrast, Suzuki et al found that vaccine effectiveness was greater in patients under 75 but this effect was not statistically significant.

### 5.5.4 Unanswered questions and future research

Although vaccination with PPV23 reduced risk of pneumococcal CAP in elderly patients, the absolute reduction in rate of disease and hospitalization was moderate. Determining an effective adult pneumococcal vaccination policy is complex because none of the available vaccines covers all serotypes and the proportion of vaccine-covered serotypes has been declining since the introduction of PCVs in children [287]. Our study was unable to look at the impact of a combined policy based on PPV23 and PCV13. In practice, the optimal adult vaccination policy will need to be flexible and adaptive, requiring monitoring of the latest available data. We showed how effectiveness of PPV23 was maintained at the oldest ages but questions remain about how the effectiveness of the vaccine varies in other population sub-groups.

### 5.5.5 Meaning of the study: possible explanations and implications for clinicians and policy makers

We found that vaccinated patients tended to be younger and experienced higher rates of comorbidity than control patients, suggesting vaccine take-up was higher

amongst patients in closer contact with the health-care system and more likely to benefit from the long-term immunity to pneumococcal disease.

The control of pneumococcal pneumonia is a public health priority in countries with an ageing population, such as the UK, because of the higher risk in older age groups and the associated health costs. Our study demonstrated a clear reduction in disease burden following the introduction of the UK policy of vaccinating older adults with PPV23. Contrary to suggestions in the literature, we found that the vaccine remained effective, and may even increase in effectiveness, at older ages, supporting the targeting of the oldest old and most frail patients for PPV23 vaccination in order to reduce the burden of pneumococcal disease. Other research would suggest vaccine efficacy declines with age in the elderly due to the age-related fall in immune response [297,298]. However, this has to be set against the increased susceptibility of the oldest age groups when assessing vaccination effectiveness in real world populations: rates of CAP increase with age in the absence of vaccination indicating the potential for benefit from immunisation may be maximised by prioritising the oldest age groups. At least in the two years after vaccination, the benefits of vaccinating the oldest adults with PPV23 to reduce susceptibility were shown to outweigh any deleterious effect of immunosenescence. These findings have implications for the formulation of future pneumococcal vaccination policy in the UK and other countries.

### 5.5.6 Conclusions

Vaccination with PPV23 has been shown to be effective in reducing risk of pneumococcal disease in patients aged 65 and older in clinical practice. The burden of disease increases with age and, despite concerns that immunosenescence may weaken immune response, we found the vaccine was most effective in the oldest old. Our study illustrates how real-world effectiveness studies with appropriate control for unmeasured confounding can provide valuable insights into the population impact of vaccination policies.

## 5.6 Appendix A

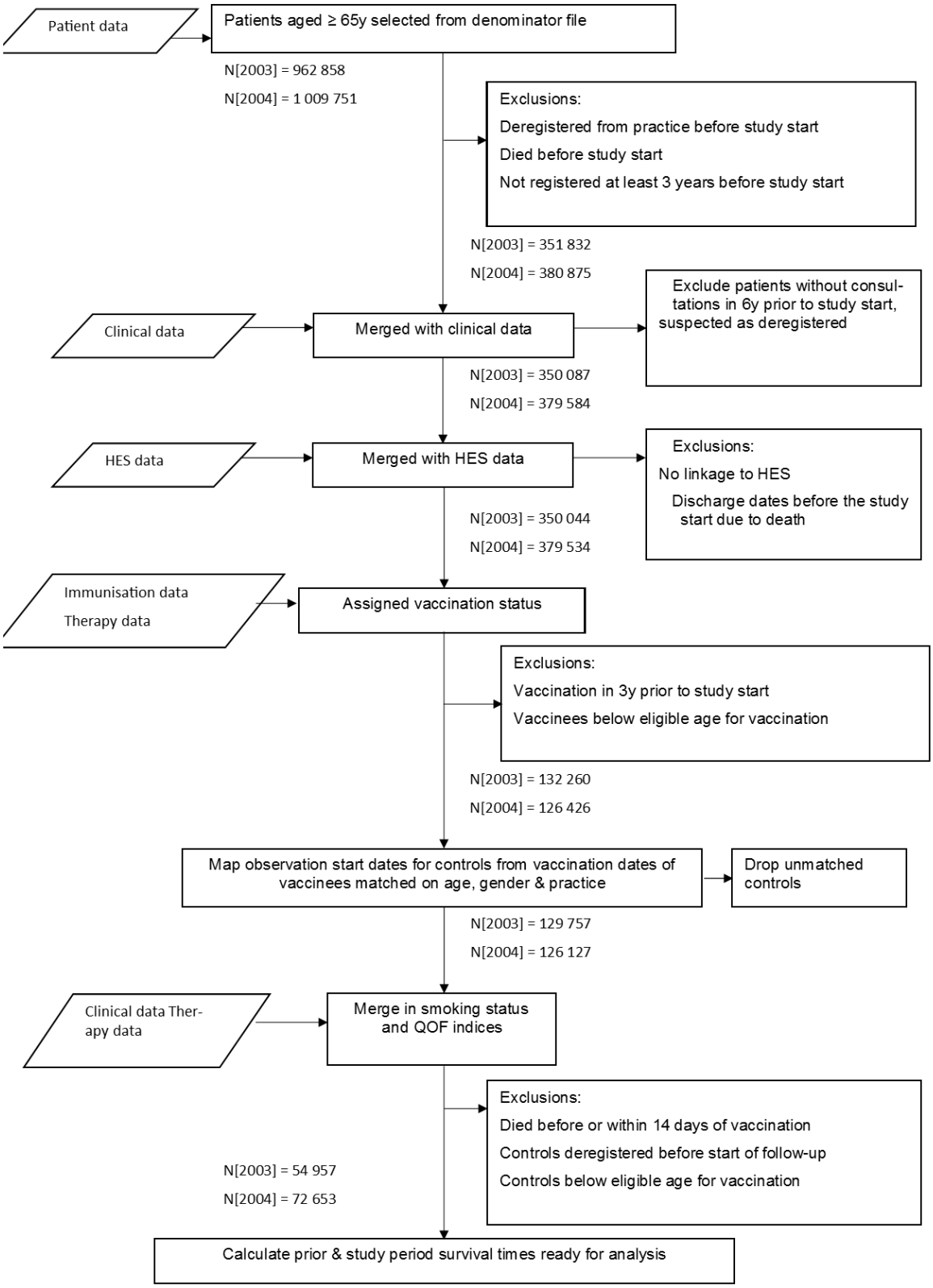


*Figure 14:* Flowchart of patient selection in study of pneumococcal vaccination

## 5.7 Appendix B: Fitting the Cox model for pneumococcal investigation

### 5.7.1 Model building

The best-fit Cox models of survival study times were found from both forwards and backwards selection. All models had to include the variable of interest: the treatment variable, trt, denoting vaccination [0,1]. The maximum level of interaction was 2-way for all interactions of the 3 main variables (gender, age, treatment). A forward selection process was used to find the best-fit model of the survival times from the prior period. The best fit was decided from the (chi-square) testing of deviance (difference in -2* log likelihood) for the additional term between two otherwise identical models. In all cases the best-fit model of the study survival times was found to apply to those from the prior period:

2003 cohort (over 79y only).

The main effect of treatment was not significantly different from the null, but the most significant model included the terms:

Main effects: Age; gender; vaccination status

Interactions: Age * gender

2004 cohort (over 74y only).

Main effects: Age; gender; vaccination status

Interactions: Age * gender

2005 cohort (over 65y)

Main effects: Age; gender; vaccination status

Interactions: Age * gender; Age * vaccination status

### 5.7.2 Diagnostic tests of PH

A test of proportional hazards through a generalised linear regression of the scaled Schoenfeld residuals on time[299], using Stata's *estat phtest*, revealed some violation of the assumption, mostly in the models applied to the prior period for study years 2003 and 2004. For those years there was evidence of non-proportionality in the test of the hazards of treatment. Given the large data size, tests were deemed significant at the 1% level. There was no evidence of non-proportionality for model of the 2005 study period, although the test of the age and the age*treatment terms in the prior period suggested some non-proportionality. With large data it is possible to detect violations that are not practically significant. However, violations in the proportionality of the hazards for the variable of interest, that of vaccination, would be of particular concern. While the test statistics for the treatment term were large enough in the prior periods of 2003 to 2004 and large enough for its interaction with age for 2005 to be flagged as statistically significant violations of the PH assumption, the regression coefficients, ρ, denoting the extent of non-proportionality were small. The negative natural logarithm of the cumulative hazard function were plotted against the natural logarithm of the analysis time to inspect the extent of any PH violation.

| | Study period | | | | | Prior period | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Term | ρ | $\chi^2$ | d.f. | Prob>χ2 | Term | ρ | $\chi^2$ | d.f. | Prob>χ2 |
| 2003 (>=80y) | gender | -0.00665 | 0.24 | 1 | 0.624 | gender | -0.00656 | 0.18 | 1 | 0.6737 |
| | age | -0.01172 | 0.74 | 1 | 0.3893 | age | 0.00726 | 0.23 | 1 | 0.6311 |
| | trt | 0.0254 | 3.64 | 1 | 0.0565 | trt | -0.07586 | 24.72 | 1 | 0 |
| | gender*age | 0.01276 | 0.88 | 1 | 0.3476 | gender*age | 0.00961 | 0.38 | 1 | 0.5383 |
| | global test | | 9.81 | 4 | 0.0437 | global test | | 28.42 | 4 | 0 |
| 2004 (>=75y) | gender | 0.02082 | 2.71 | 1 | 0.0995 | gender | 0.00909 | 0.42 | 1 | 0.5188 |
| | age | 0.01001 | 0.64 | 1 | 0.4222 | age | 0.03222 | 5.48 | 1 | 0.0193 |
| | trt | 0.02982 | 5.59 | 1 | 0.0181 | trt | -0.05709 | 16.97 | 1 | 0 |
| | gender*age | -0.01572 | 1.54 | 1 | 0.2143 | gender*age | 0.00436 | 0.1 | 1 | 0.7575 |
| | global test | | 10.42 | 4 | 0.034 | global test | | 40.47 | 4 | 0 |
| 2005 (>=65y) | gender | 0.01904 | 5.08 | 1 | 0.0242 | gender | 0.00489 | 0.3 | 1 | 0.5837 |
| | age | -0.00428 | 0.27 | 1 | 0.6049 | age | 0.02548 | 8.23 | 1 | 0.0041 |
| | trt | 0.01096 | 1.7 | 1 | 0.192 | trt | -0.00075 | 0.01 | 1 | 0.9324 |
| | gender*age | -0.00426 | 0.26 | 1 | 0.6124 | gender*age | 0.01053 | 1.42 | 1 | 0.233 |
| | age*trt | 0.00653 | 0.63 | 1 | 0.4283 | age*trt | -0.02251 | 6.69 | 1 | 0.0097 |
| | global test | | 23.76 | 5 | 0.0002 | global test | | 38.32 | 5 | 0 |

*Table 14: Table of results Stata's estat test of proportional hazards for each of the best-fit models for each year 2003-2005 (trt is the variable name for vaccination effect)*

## 5.7.3 Diagnostic plots 2003

### *5.7.3.1 Study period*

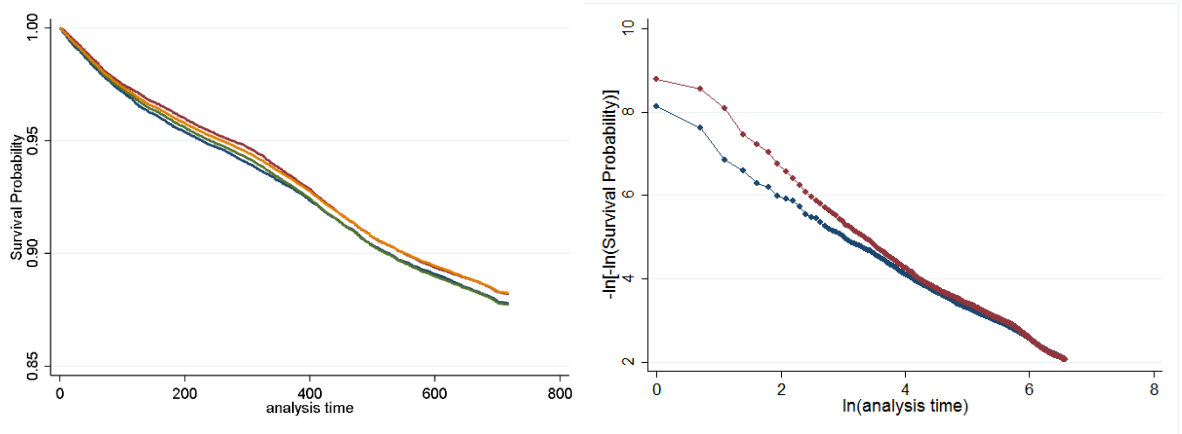


*Figure 15: Graphs of observed (vaccinated – red; controls – blue) vs. predicted (vaccinated – orange; controls – green) survival times against analysis time, and of the negative logged hazards (vaccinated – red; controls – blue) from the study period of the 2003 cohort*

While the log-log plot indicates some violation with converging lines, this would appear to occur for the relatively few short survival times (Figure 15). For the majority of the data corrresponding to the logged survival time values greater than about three (corresponding to 20 days), the hazards would appear to be reasonably proportional. The plotted predicted and observed survival curves appear reasonably close with a small descrepancy for those values corresponding to survival times of less than a year.

### *5.7.3.2 Prior period*

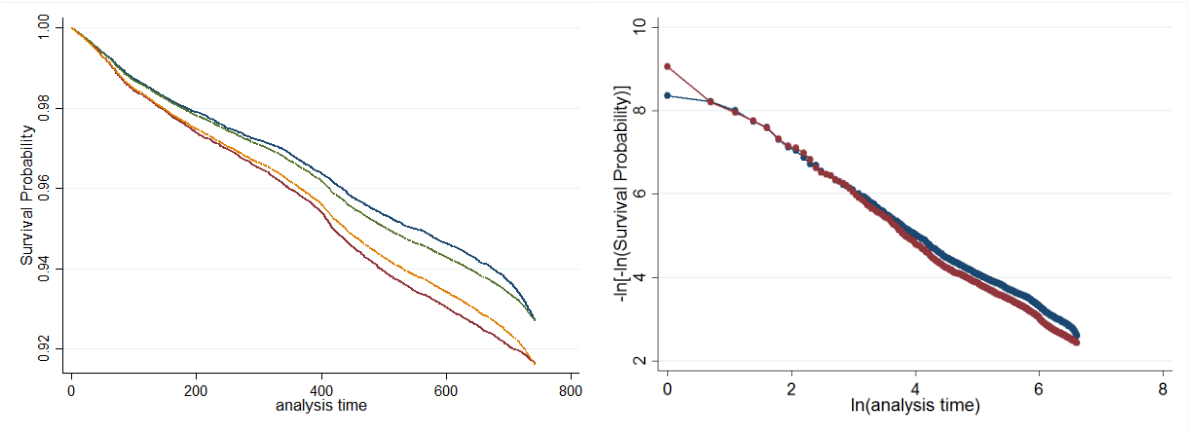


*Figure 16: Graphs of observed (vaccinated – red; controls – blue) vs. predicted (vaccinated – orange; controls – green) survival times against analysis time, and of the negative logged hazards (vaccinated – red; controls – blue) from the prior period of the 2003 cohort*

The predicted survival times from the Cox model appeared to be overestimated for the vaccines and underestimated for the controls by the second year of observation (Figure 16). The log-log plot indicated that the hazards were mostly proportional for survival times greater than about 60 days.

## 5.7.4 Diagnostic plots 2004

### *5.7.4.1 Study period*

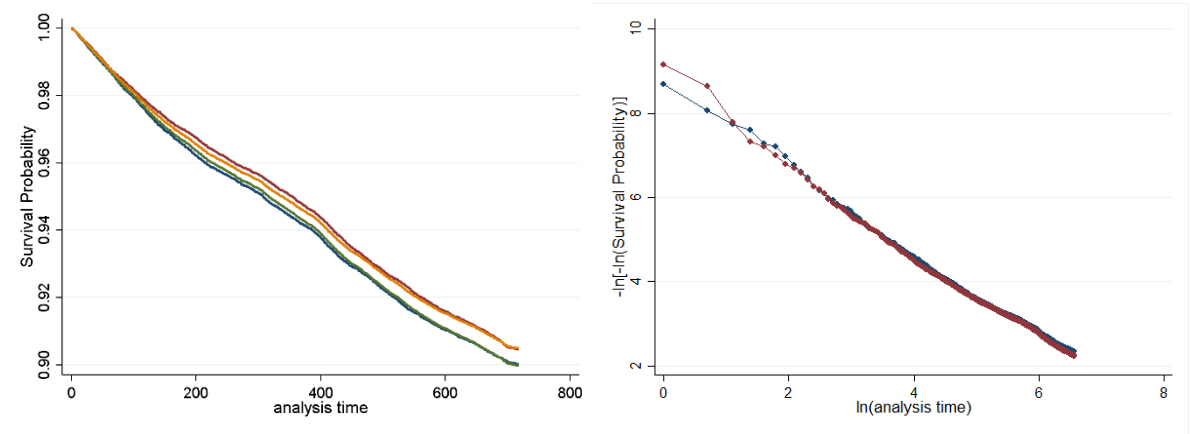


*Figure 17: Graphs of observed (vaccinated – red; controls – blue) vs. predicted (vaccinated – orange; controls – green) survival times against analysis time, and of the negative logged hazards (vaccinated – red; controls – blue) from the study period of the 2004 cohort*

Both the survival curves plot and the log-log plot indicated little if any violation of the proportional hazards assumption in the study period (Figure 17).

### *5.7.4.2 Prior period*

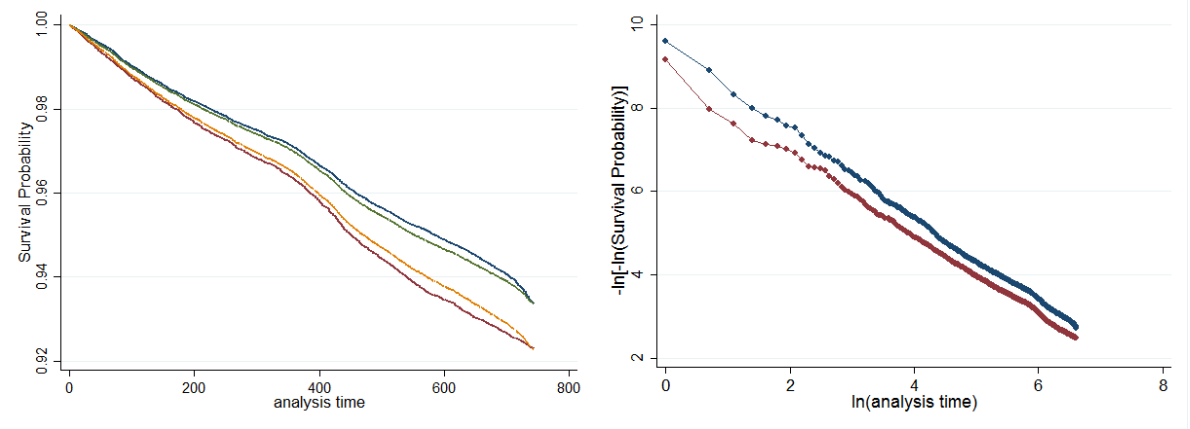


*Figure 18: Graphs of observed (vaccinated – red; controls – blue) vs. predicted (vaccinated – orange; controls – green) survival times against analysis time, and of the negative logged hazards (vaccinated – red; controls – blue) from the prior period of the 2004 cohort*

While the survival curves appeared to be overestimated for the vaccinees and underestimated for the controls, the log-log plot indicated little non-proportionality over the logged times (Figure 18).

## 5.7.5 Diagnostic plots 2005

### *5.7.5.1 Study period*

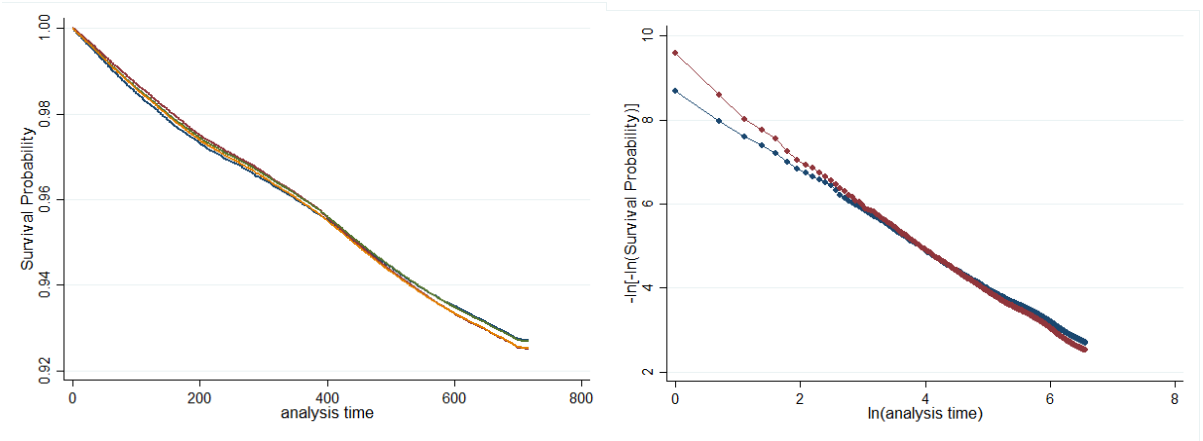


*Figure 19: Graphs of observed (vaccinated – red; controls – blue) vs. predicted (vaccinated – orange; controls – green) survival times against analysis time, and of the negative logged hazards (vaccinated – red; controls – blue) from the study period of the 2005 cohort*

From the plots it was difficult to distinguish between the predicted and observed hazards for both treatment groups , indicating little effect of treatment in the study period (Figure 19). The log-log survival plots would appear to be not entirely parallel, although small differences may be exagerrated by the plots being almost overlaid.

### *5.7.5.2 Prior period*

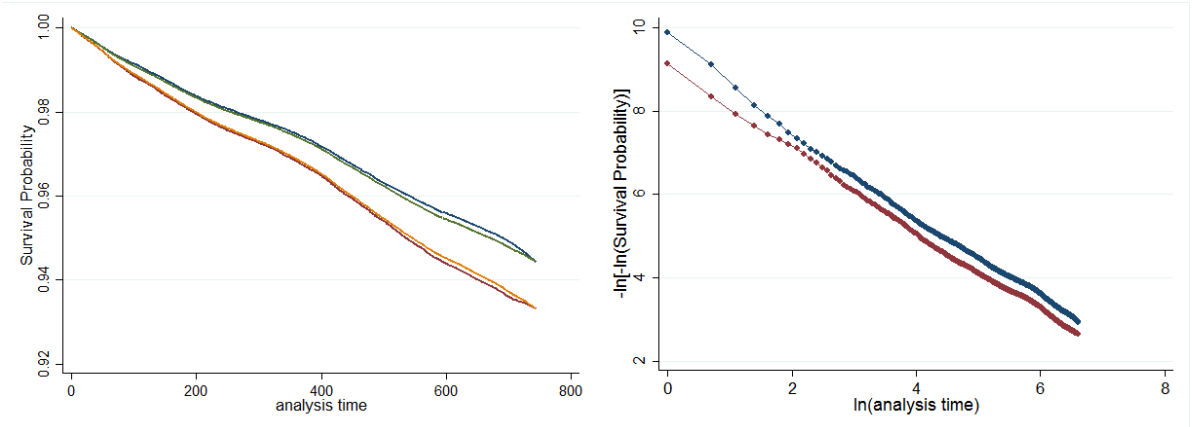


*Figure 20: Graphs of observed (vaccinated – red; controls – blue) vs. predicted (vaccinated – orange; controls – green) survival times against analysis time, and of the negative logged hazards (vaccinated – red; controls – blue) from the prior period of the 2005 cohort*

The predicted and observed survival times appear to closely agree and the log-log plots are broadly parallel (Figure 20).

### 5.7.6 Investigation of time-invariance assumption

Without prior knowledge, it is difficult to determine the optimum functional form over time of any potentially time-varying covariates, but the time-dependency of the variables was tested through the regression of survival times on the covariates, allowing for time-dependency of the coefficients on the natural logarithm of time.

| | Term varying with ln(time) | Study period | | | | Prior period | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Hazard ratio | Standard error | z | P>\|z\| | Hazard ratio | Standard error | z | P>\|z\| |
| 2003 | Treatment | 1.076268 | 0.025331 | 3.12 | 0.002 | 0.939274 | 0.027669 | -2.13 | 0.033 |
| | Age | 1.00013 | 0.002375 | 0.05 | 0.956 | 1.003576 | 0.003148 | 1.14 | 0.255 |
| | Gender | 1.044615 | 0.02522 | 1.81 | 0.071 | 1.010374 | 0.030738 | 0.34 | 0.734 |
| 2004 | Treatment | 1.042504 | 0.025705 | 1.69 | 0.091 | 0.921118 | 0.026561 | -2.85 | 0.004 |
| | Age | 0.99933 | 0.001996 | -0.34 | 0.737 | 1.005543 | 0.002554 | 2.18 | 0.029 |
| | Gender | 1.025801 | 0.024707 | 1.06 | 0.29 | 1.0536 | 0.030307 | 1.82 | 0.07 |
| 2005 | Treatment | 1.072364 | 0.018199 | 4.12 | 0 | 0.958765 | 0.017848 | -2.26 | 0.024 |
| | Age | 0.998434 | 0.001068 | -1.46 | 0.143 | 1.001681 | 0.001244 | 1.35 | 0.176 |
| | Gender | 1.034018 | 0.016828 | 2.06 | 0.04 | 1.026603 | 0.018214 | 1.48 | 0.139 |

*Table 15: Hazard ratios of the interaction between the independent variables and the natural logarithm of survival times from the models allowing for time dependency. The hazard ratios and their standard errors indicate the extent to which the hazards change over the logarithm of time from their time-invariant HRs (not shown) that were estimated in the same model.*

According to the time-varying covariate (TVC) model, there was some evidence of time dependency of the treatment effect for the study period of 2003, while for the prior period, there was little evidence of this. This was contrary to the conclusions from the regression-based test of the PH assumption (above) for the same cohort, which indicated the greatest violation of proportionality occurring in the prior rather than the study period. The situation is reversed for the 2004 cohort with significant time-dependency occurring in the prior period. For the 2005 cohort, there appeared to be a highly significant time-dependent effect of treatment in the study period, although this was not so clearly evident in the diagnostic plots of survival times. While the time-invariant HRs indicated that the direction of vaccine effectiveness with age was maintained, the time-varying estimates indicated a small convergence over time between the vaccination groups in the prior period. Conversely, in study period,

the time-varying estimates indicated a divergence over time, which could be interpreted as the waning effect of vaccination.

### 5.7.7 Conclusions

The test of proportional-hazards based on Schoenfeld residuals revealed some violations, particularly for the 2003 and 2004 cohorts, although the predicted and observed survival plots indicated that the resulting biases were likely to be small. However, the detection of such violations were powered by the size of the data. Furthermore, the results did not align with those from the investigation into the time dependency of the hazards. The PERR adjustment was applied to the time-invariant Cox models, noting the relatively minor violation of the proportional hazards assumption.

Unlike the pairwise model, the PERR adjustment very much depends on, amongst other conditions, the fit of the underlying survival model. In the case of the Cox model, proportional hazards have to be assumed. The sensitivity of the PERR method to non-proportional hazards arising from time-varying hazards was illustrated by Lin and Henley[192], which the pairwise method can avoid by specifying a time-dependent period effect. However, it would not be clear how the PERR method could be applied to survival times in the prior and study periods with differing time dependencies. Furthermore, this would considerably complicate the reporting of subsequent PERR results, as the hazards would have to be estimated for an array of time-points from the study and prior periods.

## 5.8 Appendix C: propensity score density graphs

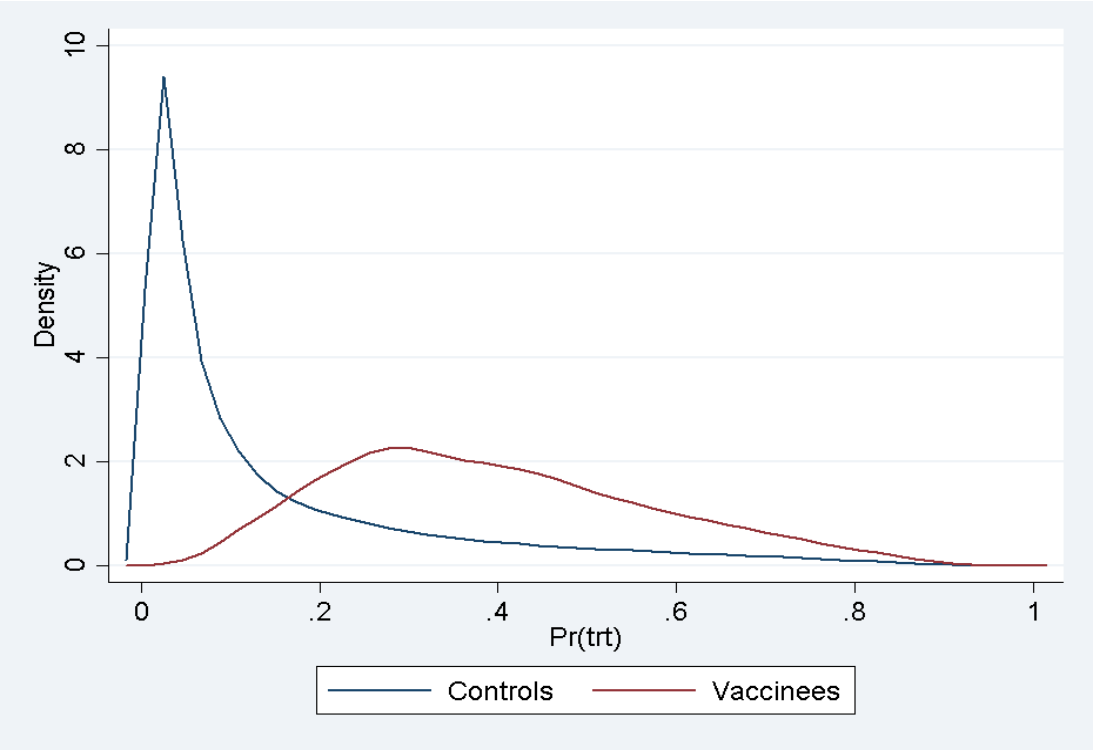


*Figure 21:* Density plot of the propensity scores for the vaccine recipients and controls in the 2003 cohort

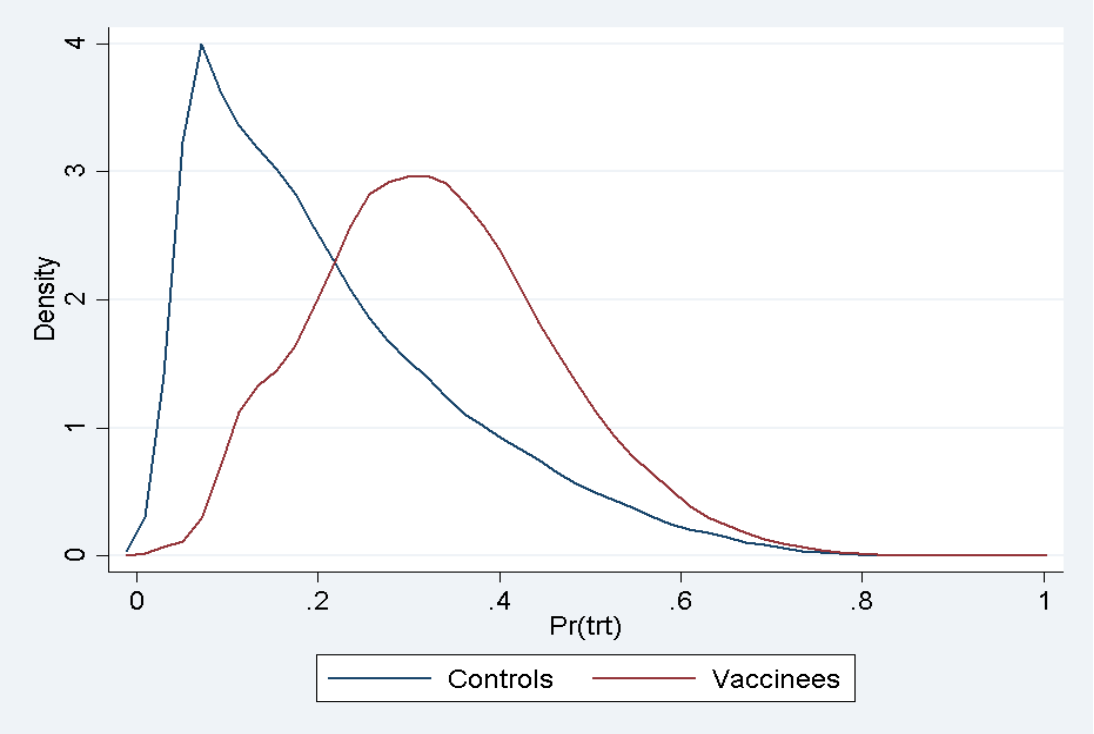


*Figure 22:* Density plot of the propensity scores for the vaccine recipients and controls in the 2004 cohort

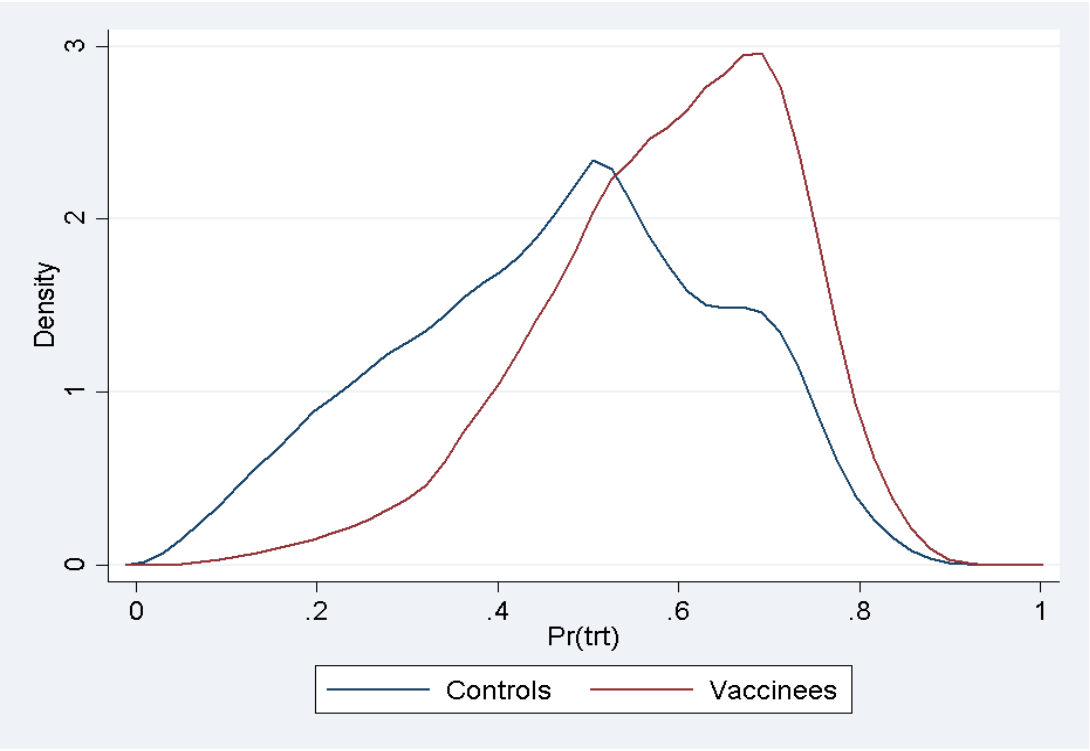


*Figure 23:* Density plot of the propensity scores for the vaccine recipients and controls in the 2005 cohort

## 5.9 Appendix D: plots of standardised mean differences of balancing variables

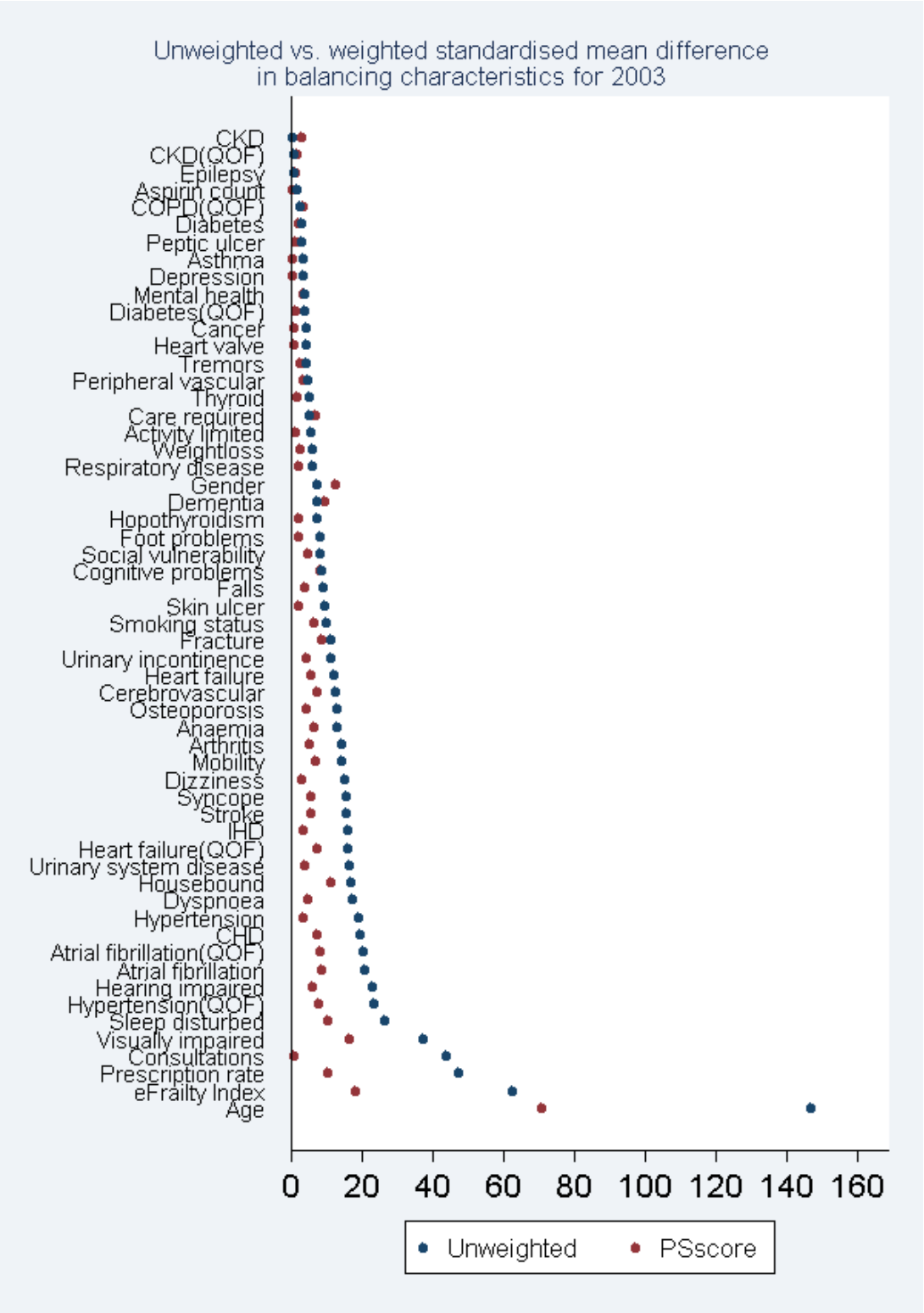


*Figure 24:* Plots of the weighted (by propensity scores) and unweighted standardised mean differences for balancing variables - 2003 cohort

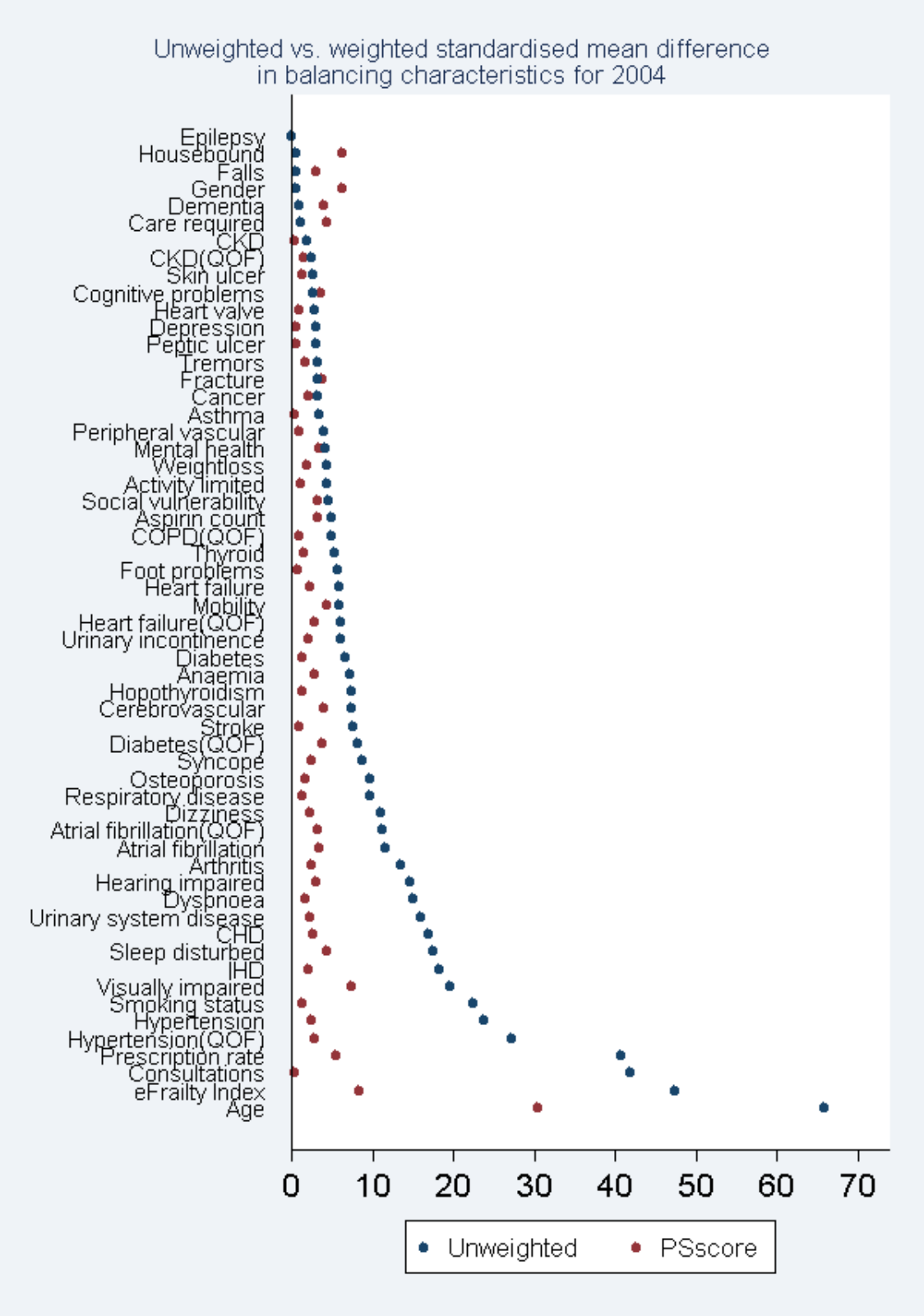


*Figure 25:* Plots of weighted (by propensity scores) and unweighted standardised mean differences for balancing variables - 2004 cohort

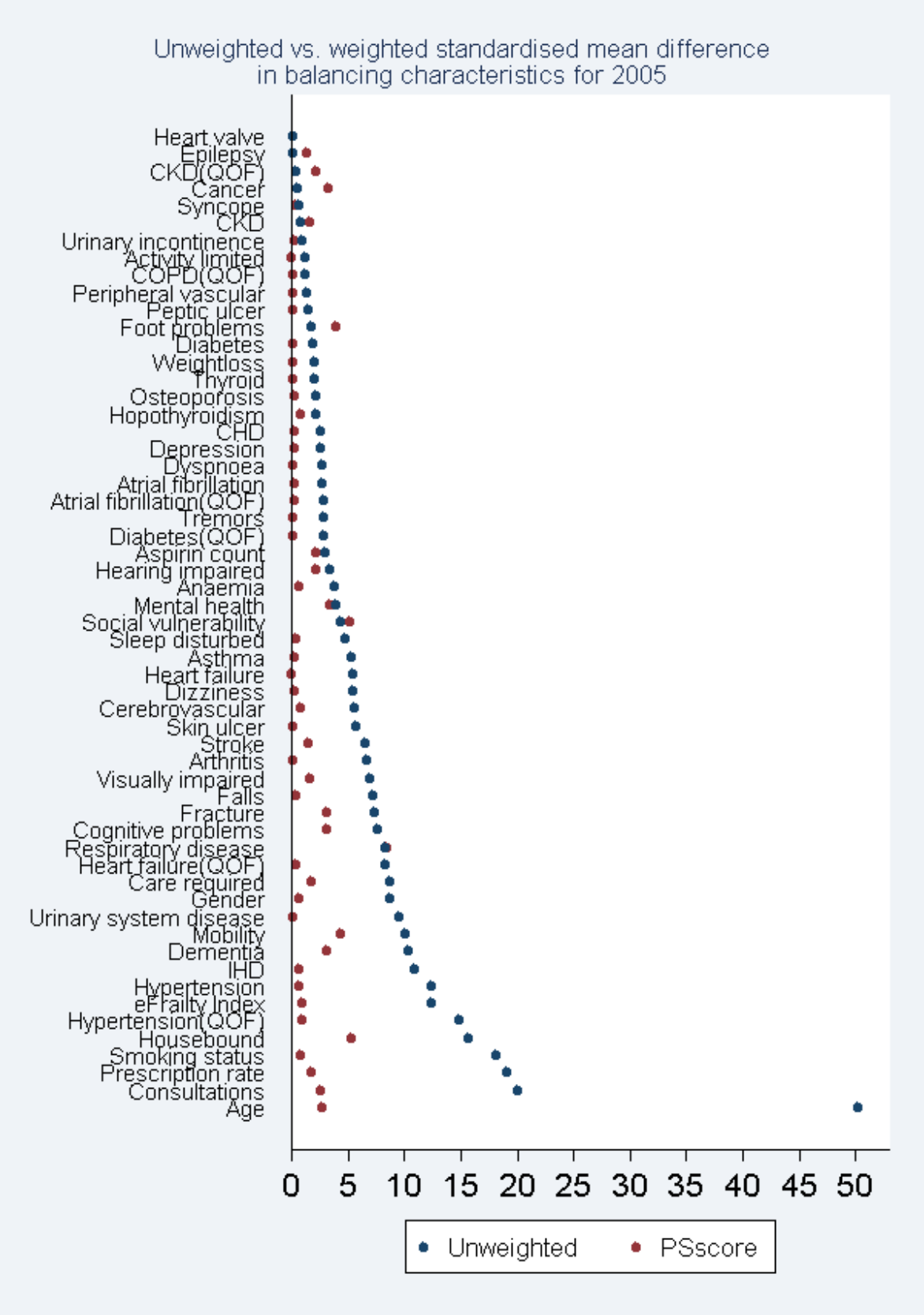


*Figure 26:* Plots of weighted (by propensity scores) and unweighted standardised mean differences for balancing variables - 2005 cohort

# Chapter 6 - Effectiveness of the influenza vaccine against myocardial infarctions in UK older adults between 1997 and 2011: a quasi-experimental cohort study

Authors:

Adam J. Streeter, Research Fellow in Medical Statistics[a,b]

Lauren Rodgers, Research Fellow in Medical Statistics[a]

Jane Masoli, NIHR Doctoral Fellow and SPR in Geriatric Medicine[c]

Alessandro Ble, Senior Research Fellow in Clinical Epidemiology[c]

Willie Hamilton, Professor of Primary Care Diagnostics[d]

David Melzer, Professor of Epidemiology and Public Health[c]

William E. Henley, Professor of Medical Statistics[a*]

[a] Health Statistics Group, University of Exeter Medical School, University of Exeter, South Cloisters, St. Luke's campus, Exeter EX1 2LU, UK

[b] Medical Satistics, Plymouth University Peninsula School of Medicine & Dentistry, University of Plymouth, Plymouth Science Park, Derriford, Plymouth PL6 8BX, UK

[c] Epidemiology & Public Health, University of Exeter Medical School, RILD Building, RD&E Hospital Wonford, Barrack Road, Exeter EX2 5DW, UK

[d] DISCOVERY research group, University of Exeter Medical School, University of Exeter, South Cloisters, St. Luke's campus, Exeter EX1 2LU, UK

## 6.1 Abstract


### 6.1.1 Objective

This study aimed to adjust for unmeasured confounding in the estimation of the real-world effectiveness of the influenza vaccine against influenza and also myocardial infarction (MI) in adults aged 65y and older in the UK.

### 6.1.2 Design

Quasi-experimental (QE) cohort study of patients in the UK from general practices registered to the Clinical Practice Research Datalink.

### 6.1.3 Setting

Adults aged 65y and over, recruited in annual cohorts, starting in September, from 1997 to 2012, with no record of a previous influenza vaccination.

### 6.1.4 Intervention

Influenza vaccination

### 6.1.5 Outcome measures

Hospitalisation for MI, influenza, and prescriptions for antibiotics for symptoms consistent with lower respiratory tract infections during a follow-up period of one year.

### 6.1.6 Results

After adjustment using the (prior event rate ratio) PERR method, the HRs for vaccination against MI from the Cox model adjusting for age and gender were significantly less than unity, varying between 0.40 (95% confidence interval: 0.30, 0.55) and 0.74 (95% CI: 0.57, 0.92), except in 2001 (HR=0.89, 95% CI: 0.68, 1.17). The same annual trend was closely mirrored in the pairwise-adjusted results, the point estimates for which varied between 0.38 and 0.80. Interaction analysis for all years apart from 1997, 1998 and 2005 indicated a waning effect of vaccination with age. The weighted estimates were greater than unity in most of the cohorts. However, upon applying the PERR adjustment, the weighted results were in accordance with those from the PERR of the adjusted Cox models. Annual differences in the mix and virulence of influenza may have given rise to the variability

between the cohorts' esimates. However, the global estimate of vaccine effectiveness across all cohorts was found to reduce the risk of MIs by 39% (95% CI: 34, 44).

Applying the same PERR method, the data demonstrated vaccination to be moderately effective against influenza in all cohorts (Pairwise HRs ranging from 0.63 in 1999 to 0.90 in 2001), although there were no significant interactions with age. Differences in effectiveness between the subgroups of pneumococcal vaccination (PPV) status were not consistent, with smaller sizes and large variability in the PPV group before 2003.

### 6.1.7 Conclusions

In spite of potential seasonal pathogen evolution and vaccine mismatches, the influenza vaccination broadly demonstrated effectiveness in older UK adults, whose annual pattern broadly followed that for effectiveness against MIs, suggesting the prevention of MIs by the vaccine may be partly mediated by influenza.

## 6.2 Introduction

The influenza vaccine is currently recommended for adults aged ≥ 65y, the age group with the highest risk of mortality from influenza viruses [300]. While vaccination is intended to protect against influenza [301], ecological data have revealed potential benefits against possible complications of the disease [302,303]. Increased hospitalisation rates for myocardial infarction (MI) and related cardiovascular conditions have long been observed to coincide with influenza epidemics [304] with an elevated risk of acute MI within seven days of laboratory-confirmed influenza infection in adults aged ≥ 65y [305]. It is thought that the influenza virus acts to increase the risk of MI both directly by provoking an inflammatory response in the heart, and indirectly by activation of inflammatory pathways and atherosclerosis, potentially exacerbated by an increased metabolic response to the virus [306,307]. The most recent Cochrane review [308] of randomised controlled trials investigating the prevention of cardiovascular events either as primary or secondary outcomes included two large studies that had investigated MI outcomes, but did not detect a significant effect [309,310]. However, the findings were based on a range of ages, and were not restricted to the elderly population.

Population-studies using electronic health records (EHR) have previously been utilised to demonstrate a protective effect of influenza vaccination, reducing the odds of MI in the year following vaccination by 20% in elderly Taiwanese patients [311]. However, EHRs are not purposed for research, and without the observations on all confounders of vaccination and MI, unmeasured confounders will likely bias an analysis that relies on adjustment for measured confounders alone. It is therefore important to diagnose and accommodate unmeasured confounding when using EHR data for inferential investigations [288].

This study therefore proposed to investigate the real-world effectiveness of influenza vaccinations against the risk of MI, as well as influenza, in the adults aged ≥ 65y registered to General Practices in UK Primary Care, using data extracted from the Clinical Practice Research Datalink. Effectiveness was examined for each annual cohort of new recipients of the influenza vaccine from 1997 to 2011, that would encompass the introduction of the policy to recommend for vaccination adults aged at least 75y in 1998, extended to those aged at least 65y in 2000. In this way, the influence of the policy itself on any trends in vaccine effectiveness could be studied. We also investigated issues with immunogenicity in this age group by analysing effectiveness with age and by sub groups of pneumococcal vaccination status.

## 6.3 Methods

### 6.3.1 Data source

The data were from the UK Clinical Practice Research Datalink (CPRD) [53], a database of electronic medical records including information on demographics, consultations, diagnoses, drug prescriptions, immunisations, referrals, etc collected by participating general practitioners during their daily clinical routines. The data analysed were from English practices, which had linkage available to hospital admissions and death certificate data. The CPRD has been granted Multiple Research Ethics Committee approval (05/MRE04/87) to undertake purely observational studies, with external data linkages including HES and ONS mortality data. The work of CPRD is also covered by NIGB-ECC approval ECC 5-05 (a) 2012. This study gained prior approval by the CPRD Independent Scientific Advisory Committee for MHRA database research (ISAC protocol 14-159).

### 6.3.2 Study population

Annual cohorts were recruited from 1997 to 2011. Recruitment began on the 1st September, the index date for each year. Patients had to be HES-linked, alive and at least 65y of age on the index date, and were excluded if not registered at their practice at least five years before. An absence of any clinical consultation in the five years before the index date was regarded as unlikely for patients in this age group and so such cases were excluded. The general practices also had to be up-to-standard at least five years before the index date.

### 6.3.3 Vaccination

Patients receiving the influenza vaccine in each annual cohort were identified according instances of influenza vaccination found under their relevant medical codes in the immunisation file and their product codes in the therapy file (Appendix D – CPRD and HES codes).

### 6.3.4 Study design and follow-up

The study comprised individual annual cohorts of eligible patients, who had not received the influenza vaccine in the two years prior to the index date. In each cohort, patients were selected for the exposed group if they had received a vaccination between the index date and the following 31st January, inclusive. Any patients receiving a vaccination during follow-up after this date were excluded from the study cohort, and any remaining unvaccinated were designated as controls for the vaccine recipients. While vaccination date determined the start of follow-up among recipients of the vaccine, obviously no such date was available for the controls. In order for there to be an approximately equal distribution of follow-up start dates among the controls as the vaccine recipients, the start dates were mapped onto the controls from the dates of vaccine recipients, either exactly matched on age, gender and GP practice, or on ever broader categories for where no exact match could be found. Matching was used solely for this purpose and played no further role in the analysis. Fourteen days were added to the vaccination dates of the recipients to allow time for full immunogenicity, and also to the start dates of the controls so that they remained commensurate with those of vaccine recipients. Any patients, for whom an MI event occurred before the attainment of full immunity, were censored and assigned a zero survival time, and so effectively did not contribute to the

likelihood of the survival model. All patients were then followed up for a year, censoring on death and transferrals out of the registered practice.

### 6.3.5 Outcomes

The primary outcome was admissions to hospital for myocardial infarction, coded according to ICD-10 (Appendix D – CPRD and HES codes), serving as the endpoint in the primary survival analysis. Where MI was not the primary reason, but an event during a subsequent episode of the same spell, admissions were not counted as endpoints. Where an admission for MI spanned the start date of either period, the start date for follow-up was adjusted to the discharge date of their hospital spell. This ensured that follow-up did not begin while already under observation in hospital for one or multiple MI events.

The secondary outcome was a composite of either hospitalisation or treatment for suspected influenza cases. Possible instances of treatment were prescriptions for antiviral drugs used to treat influenza, or lower respiratory tracts infections requiring treatment with antibiotics (Appendix D – CPRD and HES codes). Antibiotic prescriptions were identified in the data by their corresponding codes, and qualified by codes for symptoms that had previously been validated by two clinicians as being consistent with those for lower respiratory tract infections (Appendix D – CPRD and HES codes). Hospitalisations for influenza were indicated in the HES data by their corresponding ICD10 codes describing the reason for admission.

### 6.3.6 Statistical analysis

The effect of influenza vaccination on survival times until the first MI was analysed using Cox's regression adjusting for age (centred on 65y) and gender, censoring on death. Any patients found to have vaccination dates occurring after their date of death were dropped from the cohort. Any negative survival times resulting from the addition of 14 days (reflecting the period for immunogenicity) to the vaccination date were assigned zero times, with the corresponding events coded as right-censored events. As a guide to possible effect sizes, results for each annual cohort were presented with 95% confidence intervals, and so inference was carried out at the 5% significance level. The same approach was used for the secondary outcome of influenza events.

As an alternative approach to modelling time and understanding the effect of vaccination relative to a common index date, the data were also expressed as a counting process for analysis of vaccination as a time-varying covariate (TVC). By initiating follow-up from a common index date of 1st September, this simplified the analysis, but also could account for the time at risk in the vaccine recipients before their vaccination. Patients in the intervention group, therefore, had two possible survival times in the study period and were coded as vaccine-free up until vaccination, but vaccinated thereafter. In doing so, data preparation was greatly simplified, and the analysis could be run without the need for matching and assigning controls a start date based on vaccination dates. Follow-up for those patients, who were still in hospital from a previous MI at the time of the index date, began from the date of discharge. For the vaccine recipients, follow-up re-continued from the vaccination date plus 14 days, with the vaccination indicator variable set to one.

#### *6.3.6.1 Adjustment for confounding*

An adjustment on measured confounders available in the data was attempted through an inverse probability treatment weighted (IPTW) analysis of each individual cohort. For this, the propensity score summarising the probability of vaccination for every patient had to be derived for each cohort from a logistic regression model of vaccination status on gender, age, indices of the Quality Outcomes Framework [295] and deficits in the electronic Frailty Index (eFI) [292], as well as the eFI itself . To manage the size of the task, and avoid individually fitting a model for every cohort, a program was written in Stata to fit an optimum model for each cohort according to the significance of the Wald statistics of each regressor at the 5% level.

To mitigate for confounding bias, both measured and hidden, the prior event rate ratio (PERR) method [242,243] was applied to the survival data. For comparison, confounding was also adjusted through an alternative approach to the PERR, applying the likelihood framework of the Pairwise method [192].

The PERR method was applied to the data presented in the counting process format, modelling vaccination state as a time-varying coefficient [278]. Adjustment comes from dividing the estimate for the vaccination effect in the study period by that from the vaccine-free prior period. This required following up the patients from the date one-year prior to their start date in the study period, until either the defined MI

event or the index date of the study period. The PERR method was applied to the Cox regression estimates for vaccine effectiveness and also to those from the TVC model for each annual cohort. In both applications, the format of data from the prior period was the same, as no vaccinations occurred in the prior period, and therefore each patient had only one survival time.

The PERR method was also applied to the IPTW results. In order to derive the weights for the prior period, the automated logistic model selection was run for the prior periods to each cohort, as done for the study period.

As well as analysing each individual cohort, the data from all cohorts were aggregated into a single, global dataset, and analysed using PERR-adjusted Cox models adjusting for age and gender. Because this approach would likely aggregate the data on the same patients from across several cohorts, the same patients could be represented across several cohorts. To account for a potential lack of independence between observations, these were analysed using robust standard errors, clustered on patient id.

#### *6.3.6.2 Sub-group analysis*

Further analysis tested for any moderating effect of age, by modelling the interaction between age and vaccination status and their main effects, to which the PERR adjustment was applied. In a sensitivity analysis of the effectiveness of the influenza vaccine was re-analysed according to polysaccharide pneumococcal vaccination (PPV) status. Both the PERR and pairwise analyses were repeated for subgroups of PPV, classified into patient, who have had a record of PPV in both study and the prior period (ever), and those with no record of PPV (never).

## 6.4 Results

### 6.4.1 Cohort characteristics

There was an overall increase in the size of the annual cohorts from 62 644 in 1997 to 130 460 in 2011, while the annual percentage rate of influenza vaccinations among the 65+ year old patients of each cohort fluctuated around 15% from 1997 to 1999 (Table 16). However, the rate increased to 39.5% in 2000 with the introduction of the policy to increase vaccine coverage in the adults aged at least 65y. The

exclusion of patients with an influenza vaccination in the two years prior to recruitment likely contributed to the reduction of uptake in vaccination in the following cohorts to a minimum of 56370 in 2002. Thereafter the cohort size steadily increased, while the vaccination rate fluctuated between 12.9% in 2007 and 24.5% in 2005.

The mean age of each cohort remained at around 74y until 2003, decreasing very slightly to below 73y by 2010. While vaccine recipients were the same age or slightly older than the controls up until 2000, their mean age decreased with each year to about 70y in 2011, eventually about 3.4y younger than the controls. Patients with at least one QOF-registered disease comprised 68.2% of the vaccine recipients compared to 54.1% of the controls in the 1997 cohort, but by 2011 this disparity had steadily reduced to 60.6% and 58.1%, respectively. This trend in disparity may have been driven at least by the most commonly diagnosed condition, hypertension, the prevalence of which in the vaccine recipients and controls stood, respectively, at 32.6% and 27.3% in 1997 increasing to similar levels, 36.3% and 35.3%, by 2011. The difference between vaccination groups in the prevalence of the next most frequent morbidity, coronary heart disease, also narrowed from 19.3% and 13.4%, in vaccine recipients and controls respectively in 1997, to similar levels, 7.3% and 8.8%, in 2011. Similar declining trends were seen in atrial fibrillation, asthma, chronic obstructive pulmonary disease, depression and strokes. The remaining diseases had relatively low levels, except for chronic kidney disease, which increased from no recorded diagnoses before 2001 to 14.2% of the controls and 11.3% of vaccine recipients in 2011.

| Year | N | N vaccinated | % vaccinated | Vaccine status | N MI admissions | Mean age | % males | % patients with QOF diseases | Atrial fibrillation | Asthma | Cancer | CHD | CKD | COPD | Dementia | Depression | Diabetes | Epilepsy | HF | Hyper-tension | Hypo-thyroid | Mental health | Stroke |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1997 | 62644 | 7687 | 12.3 | Controls | 369 | 74.1 | 41.9 | 54.1 | 4.4 | 5.7 | 4.6 | 13.4 | 0 | 3.9 | 1.4 | 8.7 | 1.2 | 0.9 | 5.3 | 27.3 | 3.9 | 0.8 | 6.7 |
| | | | | Vaccinated | 71 | 74.3 | 41.9 | 68.2 | 6.4 | 10.4 | 5.4 | 19.3 | 0 | 7.5 | 2 | 11.6 | 2.5 | 1.1 | 8 | 32.6 | 5.4 | 0.7 | 9.2 |
| 1998 | 68421 | 5801 | 8.5 | Controls | 431 | 74 | 42.3 | 54.4 | 4.4 | 6 | 4.6 | 13.6 | 0 | 3.9 | 1.3 | 8.2 | 1.4 | 0.9 | 4.9 | 28.3 | 4.1 | 0.8 | 6.4 |
| | | | | Vaccinated | 55 | 76.1 | 42.3 | 71.6 | 7.5 | 10.3 | 6 | 21.3 | 0 | 7.3 | 3.2 | 11.7 | 3.6 | 1.3 | 9.3 | 35.8 | 5.2 | 0.7 | 9.8 |
| 1999 | 72288 | 8686 | 12 | Controls | 437 | 73.9 | 42.8 | 54.3 | 4.3 | 5.8 | 4.4 | 12.9 | 0 | 3.6 | 1.2 | 7.9 | 1.7 | 0.9 | 4.4 | 29.4 | 4.2 | 0.8 | 5.9 |
| | | | | Vaccinated | 102 | 75.3 | 42.8 | 70.2 | 7.3 | 10.3 | 5.6 | 21.3 | 0 | 7.7 | 2.3 | 11.2 | 3.4 | 0.9 | 7.8 | 36.3 | 5.5 | 0.9 | 9 |
| 2000 | 73527 | 29058 | 39.5 | Controls | 356 | 74.5 | 42.8 | 53.5 | 4.5 | 5.6 | 4.4 | 12.4 | 0 | 3.5 | 1.5 | 7.5 | 2 | 0.9 | 4.3 | 29.2 | 4.3 | 0.9 | 5.9 |
| | | | | Vaccinated | 250 | 72.9 | 42.8 | 61.7 | 5 | 7 | 5.1 | 15.3 | 0 | 4.1 | 1.2 | 8.3 | 2.9 | 1 | 4.2 | 35.4 | 5 | 0.7 | 6.2 |
| 2001 | 58998 | 13753 | 23.3 | Controls | 337 | 74.6 | 41.6 | 53.1 | 4.3 | 5.6 | 4.3 | 12 | 0 | 3.5 | 1.4 | 7.2 | 2.4 | 0.9 | 4 | 29.6 | 4.7 | 0.9 | 5.7 |
| | | | | Vaccinated | 146 | 73.2 | 41.6 | 62.6 | 5.4 | 6.8 | 4.9 | 15.5 | 0.1 | 4.3 | 1.6 | 8.1 | 3.9 | 1.1 | 4.4 | 35.6 | 5.7 | 0.8 | 6.7 |
| 2002 | 56370 | 9875 | 17.5 | Controls | 361 | 74.4 | 41.7 | 52.6 | 4.4 | 5.4 | 4.3 | 11.5 | 0.1 | 3.2 | 1.5 | 6.9 | 3 | 0.9 | 3.6 | 30 | 5 | 0.9 | 5.4 |
| | | | | Vaccinated | 95 | 71.9 | 41.7 | 63.2 | 4.7 | 6.7 | 5.8 | 14.1 | 0.1 | 4.1 | 2 | 8.4 | 5 | 1.2 | 3.8 | 36.8 | 5.9 | 0.8 | 6.1 |
| 2003 | 59851 | 10943 | 18.3 | Controls | 395 | 74.4 | 42.1 | 53.5 | 4.6 | 5.5 | 4.5 | 11.1 | 0.1 | 3.2 | 1.6 | 6.6 | 3.9 | 0.9 | 3.1 | 31.5 | 5.2 | 0.9 | 5.3 |
| | | | | Vaccinated | 73 | 71.5 | 42.1 | 62.1 | 5 | 6.7 | 5 | 12.6 | 0.1 | 4.4 | 2.1 | 8.5 | 5.6 | 1.2 | 3 | 37 | 6.2 | 0.9 | 5.6 |
| 2004 | 69285 | 11896 | 17.2 | Controls | 379 | 74.2 | 42.7 | 53.6 | 4.6 | 5.4 | 4.6 | 10.4 | 0.3 | 3.1 | 1.5 | 6.1 | 4.8 | 0.9 | 2.7 | 32.5 | 5.6 | 0.9 | 5 |
| | | | | Vaccinated | 90 | 71.3 | 42.7 | 65.4 | 5.2 | 7.5 | 5.6 | 12.9 | 0.3 | 5 | 2.2 | 7.8 | 7.1 | 1.1 | 3 | 39.6 | 6 | 1.1 | 6.1 |
| 2005 | 81591 | 20027 | 24.5 | Controls | 406 | 74.4 | 42.8 | 53.9 | 4.9 | 5.4 | 4.7 | 10.1 | 3.2 | 3.3 | 1.6 | 5.6 | 5.6 | 0.9 | 2.5 | 32.7 | 5.9 | 1 | 5.1 |
| | | | | Vaccinated | 132 | 71 | 42.8 | 60.3 | 4.3 | 5.8 | 5.8 | 10.1 | 3.2 | 3.8 | 1.8 | 6.9 | 6.4 | 1.1 | 1.9 | 36.9 | 6.3 | 0.9 | 4.6 |
| 2006 | 77136 | 10635 | 13.8 | Controls | 410 | 74.1 | 43.2 | 54.9 | 4.9 | 5.3 | 4.8 | 9.8 | 9.6 | 3.3 | 1.6 | 5.7 | 6.3 | 0.8 | 2.1 | 33.4 | 6.2 | 1.2 | 4.7 |
| | | | | Vaccinated | 68 | 70.6 | 43.2 | 61.3 | 5.1 | 4.9 | 5.6 | 9.8 | 10.4 | 4.4 | 2.6 | 7 | 6.7 | 1.2 | 2.3 | 36.8 | 6.4 | 1.1 | 5.2 |
| 2007 | 87388 | 11286 | 12.9 | Controls | 460 | 73.8 | 44.1 | 55.1 | 5 | 5.2 | 5 | 9.2 | 11.6 | 3.3 | 1.6 | 5.4 | 6.4 | 0.8 | 1.8 | 33.7 | 6.3 | 1.2 | 4.4 |
| | | | | Vaccinated | 65 | 69.9 | 44.1 | 61.6 | 4.6 | 5.1 | 6 | 8.9 | 11.2 | 4 | 2.6 | 7.5 | 6.6 | 1 | 1.8 | 36.9 | 6.7 | 1.4 | 4.3 |
| 2008 | 97355 | 16225 | 16.7 | Controls | 555 | 73.6 | 44.4 | 55.9 | 5 | 5.2 | 5 | 9 | 12.7 | 3.4 | 1.6 | 5.4 | 6.8 | 0.8 | 1.7 | 34.2 | 6.5 | 1.2 | 4.2 |
| | | | | Vaccinated | 90 | 69.7 | 44.4 | 60.7 | 4.5 | 5 | 6.2 | 7.6 | 11 | 3.7 | 2.2 | 6.9 | 6.3 | 1 | 1.4 | 36.5 | 7 | 1.3 | 3.5 |
| 2009 | 103538 | 14839 | 14.3 | Controls | 537 | 73.3 | 44.8 | 56.1 | 5 | 5 | 4.9 | 8.6 | 12.8 | 3.5 | 1.5 | 5.2 | 7 | 0.8 | 1.5 | 34.6 | 6.5 | 1.1 | 4 |
| | | | | Vaccinated | 93 | 69.6 | 44.8 | 61.5 | 4.3 | 5.6 | 5.9 | 7.7 | 12.1 | 4.2 | 2.2 | 6.7 | 6.5 | 1.1 | 1.4 | 36.9 | 6.8 | 1.1 | 3.8 |
| 2010 | 113666 | 15197 | 13.4 | Controls | 601 | 73.5 | 45 | 57.3 | 5.1 | 5.2 | 4.7 | 8.7 | 13.7 | 3.7 | 1.6 | 4.9 | 7.6 | 0.8 | 1.4 | 35.2 | 6.8 | 1.1 | 3.7 |
| | | | | Vaccinated | 66 | 69.7 | 45 | 61.2 | 4.9 | 5.3 | 6.2 | 7.8 | 11.4 | 4.4 | 2.6 | 6.4 | 7.2 | 1.2 | 1.3 | 36.3 | 7.2 | 1.3 | 3.8 |
| 2011 | 130460 | 20302 | 15.6 | Controls | 696 | 73.4 | 45.2 | 58.1 | 5.2 | 5.5 | 4.3 | 8.8 | 14.2 | 4 | 1.6 | 4.4 | 8.2 | 0.8 | 1.3 | 35.3 | 7 | 1.1 | 3.4 |
| | | | | Vaccinated | 101 | 70 | 45.2 | 60.6 | 4.9 | 4.9 | 5.7 | 7.3 | 11.3 | 4.3 | 2.4 | 5.9 | 6.7 | 0.9 | 1.3 | 36.3 | 7.3 | 1 | 3.2 |

*Table 16:* Table of each annual cohort's characteristics describing vaccination status, hospital admissions for myocardial infarctions, age, gender and proportions of diseases monitored under the Quality Outcomes Framework.

## 6.4.2 Effectiveness of influenza vaccination on MI

### *6.4.2.1 Cox models and PERR adjustment*

The Cox model in both the study and prior periods of every cohort were adjusted for age and gender (diagnostic log-log plots for the fitted Cox models in Appendix E). The hazard ratios (HRs) for gender in every cohort were significantly greater than unity, varying between 1.40 and 1.87, indicating a greater risk of MI in males compared to the reference level of females (Table 17). A greater than unity HR for the age variable indicated increasing risks with age too, that were statistically significant. Effects in the same direction and of similar size were seen in the prior period too for the age and gender variables of each cohort's Cox model.

The HR for vaccination was greater than unity in the study period of every cohort apart from 2003, 2008 and 2010. These were significantly different from unity in years 1997, 1999, 2001, 2002 and 2004. However, the HRs for the study-period vaccination status in the vaccine-free prior periods were all greater than unity and greater than the study HRs, suggesting the presence of a pre-existing confounding bias (Figure 27). There was no discernible similarity in the trends of the HRs over time of the prior and study periods. However, there was noticeably more variability of the prior-period point estimates, ranging from 1.43 to 2.67, (orange circles in Figure 27) compared to those for the study period, ranging from 0.93 to 1.57 (blue circles in Figure 27). With fewer events in the prior, the confidence intervals were wider too. After the PERR adjustment, the estimated HR of influenza vaccination varied mostly between 0.40 in 2010 and 0.74 in 2000 and 2002 (Table 17; Figure 28). All PERR-adjusted estimates were all significantly different from unity, except for 2001, for which the HR was 0.89 (95% CI: 0.70, 1.17). Therefore, the results would indicate a reduction in the risk of MI by as much as 60% might be possible following an influenza vaccination.

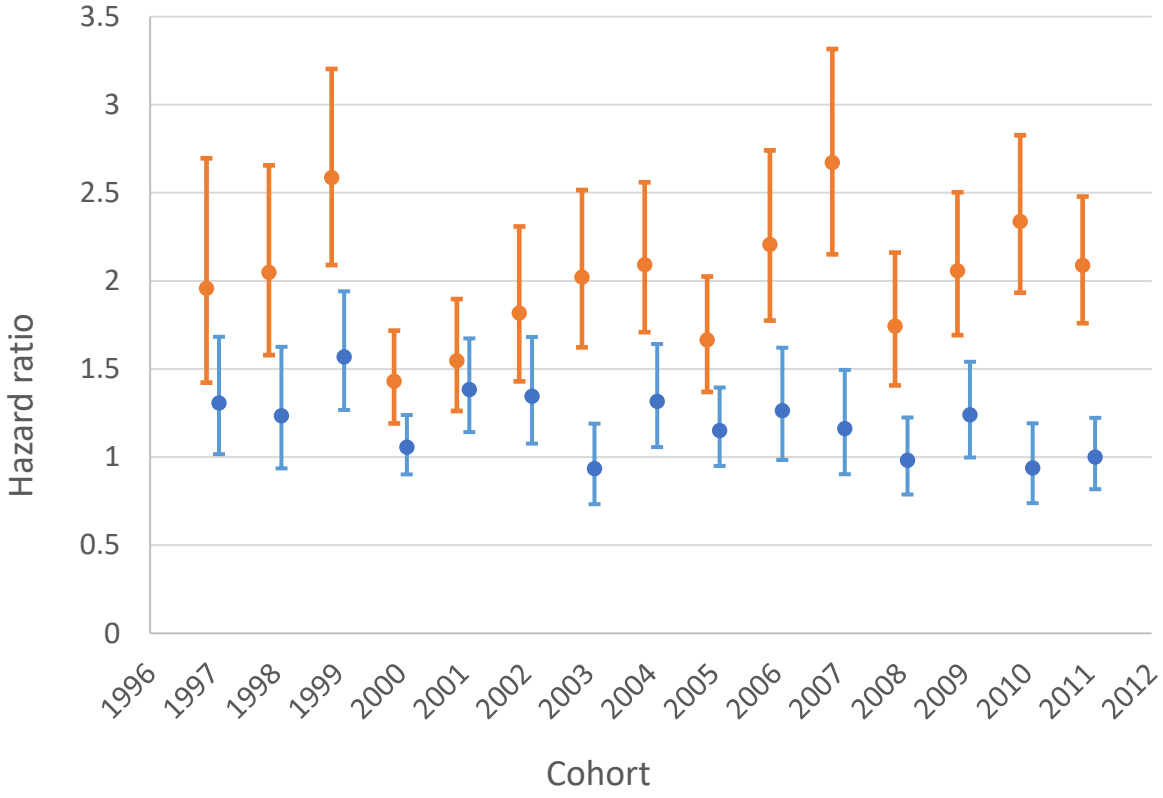


*Figure 27:* Plot of hazard ratios of the effect of influenza vaccination on MIs from the Cox regression models adjusted for age and gender for the prior (orange) and study (blue) periods of each annual cohort with error bars representing the bootstrapped 95% confidence intervals.

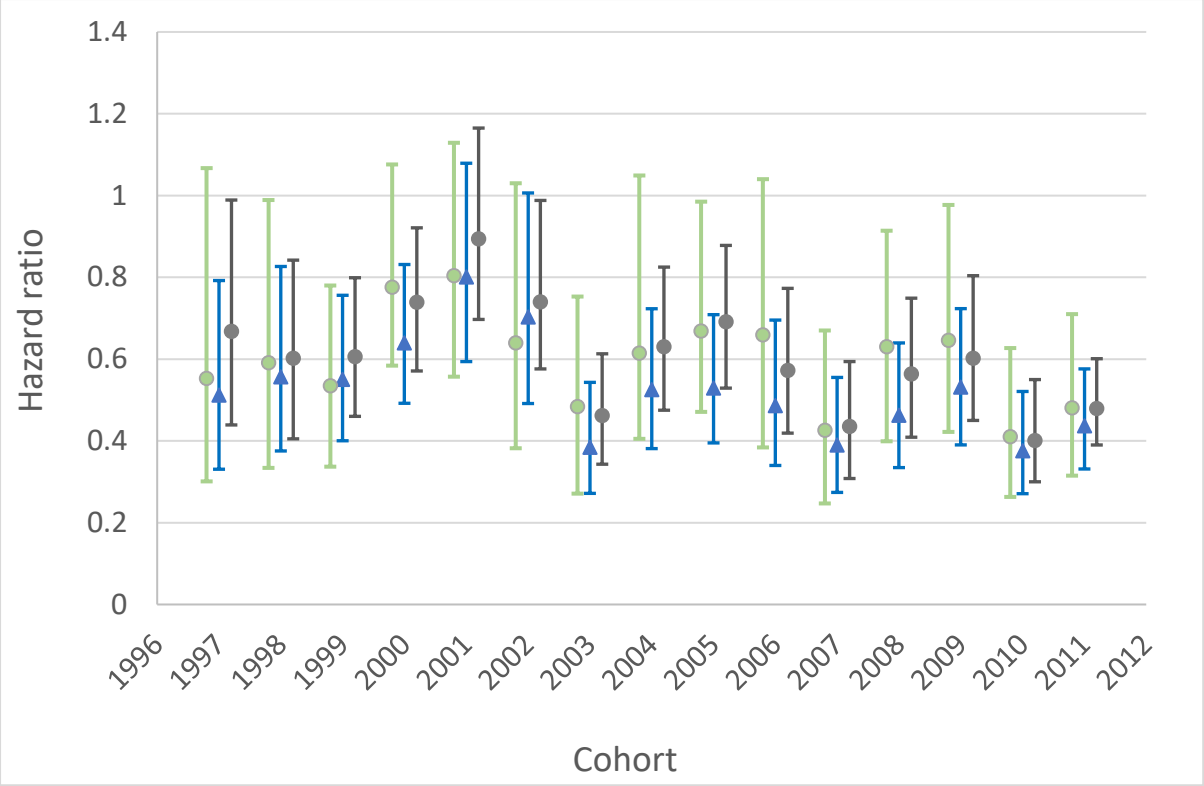


*Figure 28:* Plot of hazard ratios for the estimated effect of influenza vaccination on MI hospital admissions from the PERR-adjusted model (grey dots), the pairwise model (blue triangles) and the PERR-adjusted time-varying covariate model (green dots) with errors bars representing the bootstrapped 95% confidence intervals.

| Recruityear | Period | No of MI | | HR gender | HR age | HR FV group |
|---|---|---|---|---|---|---|
| | | Vaccinated | Controls | | | |
| 1997 | Prior | 48 | 174 | 1.61 (1.24, 2.11) | 1.01 (1.00, 1.03) | 1.96 (1.42, 2.70) |
| 1997 | Study | 72 | 386 | 1.55 (1.29, 1.87) | 1.03 (1.02, 1.04) | 1.31 (1.02, 1.68) |
| 1997 | PERR | | | | | 0.67 (0.44, 0.99) |
| 1998 | Prior | 69 | 342 | 1.82 (1.50, 2.22) | 1.03 (1.02, 1.04) | 2.05 (1.58, 2.66) |
| 1998 | Study | 57 | 461 | 1.75 (1.47, 2.08) | 1.04 (1.03, 1.05) | 1.23 (0.94, 1.63) |
| 1998 | PERR | | | | | 0.60 (0.41, 0.84) |
| 1999 | Prior | 116 | 319 | 1.51 (1.25, 1.83) | 1.02 (1.01, 1.03) | 2.59 (2.09, 3.20) |
| 1999 | Study | 105 | 452 | 1.87 (1.58, 2.21) | 1.04 (1.03, 1.05) | 1.57 (1.27, 1.94) |
| 1999 | PERR | | | | | 0.61 (0.46, 0.80) |
| 2000 | Prior | 224 | 244 | 1.87 (1.55, 2.25) | 1.03 (1.02, 1.04) | 1.43 (1.19, 1.72) |
| 2000 | Study | 260 | 382 | 1.84 (1.57, 2.15) | 1.04 (1.03, 1.06) | 1.06 (0.90, 1.24) |
| 2000 | PERR | | | | | 0.74 (0.57, 0.92) |
| 2001 | Prior | 136 | 296 | 2.02 (1.67, 2.46) | 1.04 (1.03, 1.05) | 1.55 (1.26, 1.90) |
| 2001 | Study | 149 | 365 | 1.84 (1.54, 2.19) | 1.05 (1.04, 1.06) | 1.38 (1.14, 1.67) |
| 2001 | PERR | | | | | 0.89 (0.70, 1.17) |
| 2002 | Prior | 92 | 263 | 1.61 (1.30, 1.99) | 1.05 (1.04, 1.06) | 1.90 (1.49, 2.41) |
| 2002 | Study | 99 | 382 | 1.70 (1.42, 2.04) | 1.06 (1.05, 1.07) | 1.31 (1.04, 1.64) |
| 2002 | PERR | | | | | 0.74 (0.58, 0.99) |
| 2003 | Prior | 114 | 292 | 1.45 (1.19, 1.77) | 1.06 (1.04, 1.07) | 2.02 (1.62, 2.52) |
| 2003 | Study | 79 | 421 | 1.73 (1.44, 2.07) | 1.05 (1.04, 1.07) | 0.93 (0.73, 1.19) |
| 2003 | PERR | | | | | 0.46 (0.34, 0.61) |
| 2004 | Prior | 134 | 347 | 1.63 (1.36, 1.96) | 1.05 (1.04, 1.06) | 2.09 (1.71, 2.56) |
| 2004 | Study | 100 | 417 | 1.67 (1.40, 1.99) | 1.07 (1.06, 1.08) | 1.32 (1.06, 1.64) |
| 2004 | PERR | | | | | 0.63 (0.48, 0.83) |
| 2005 | Prior | 154 | 334 | 1.48 (1.24, 1.78) | 1.05 (1.04, 1.06) | 1.67 (1.37, 2.03) |
| 2005 | Study | 141 | 456 | 1.53 (1.30, 1.80) | 1.07 (1.06, 1.08) | 1.15 (0.95, 1.40) |
| 2005 | PERR | | | | | 0.69 (0.53, 0.88) |
| 2006 | Prior | 108 | 367 | 1.56 (1.30, 1.88) | 1.06 (1.05, 1.07) | 2.21 (1.78, 2.74) |
| 2006 | Study | 73 | 449 | 1.73 (1.45, 2.07) | 1.08 (1.07, 1.09) | 1.26 (0.98, 1.62) |
| 2006 | PERR | | | | | 0.57 (0.42, 0.77) |
| 2007 | Prior | 112 | 353 | 1.64 (1.36, 1.98) | 1.06 (1.05, 1.07) | 2.67 (2.15, 3.32) |
| 2007 | Study | 70 | 516 | 1.70 (1.44, 2.00) | 1.07 (1.06, 1.08) | 1.16 (0.90, 1.50) |
| 2007 | PERR | | | | | 0.44 (0.31, 0.59) |
| 2008 | Prior | 111 | 401 | 1.52 (1.27, 1.82) | 1.06 (1.05, 1.07) | 1.74 (1.41, 2.16) |
| 2008 | Study | 94 | 606 | 1.48 (1.27, 1.73) | 1.07 (1.06, 1.08) | 0.98 (0.79, 1.23) |
| 2008 | PERR | | | | | 0.56 (0.41, 0.75) |
| 2009 | Prior | 133 | 476 | 1.42 (1.21, 1.67) | 1.06 (1.05, 1.07) | 2.06 (1.69, 2.50) |
| 2009 | Study | 97 | 597 | 1.40 (1.20, 1.63) | 1.07 (1.06, 1.08) | 1.24 (1.00, 1.54) |
| 2009 | PERR | | | | | 0.60 (0.45, 0.80) |
| 2010 | Prior | 141 | 500 | 1.42 (1.21, 1.67) | 1.07 (1.06, 1.08) | 2.34 (1.93, 2.83) |
| 2010 | Study | 76 | 678 | 1.62 (1.40, 1.88) | 1.08 (1.07, 1.09) | 0.94 (0.74, 1.19) |
| 2010 | PERR | | | | | 0.40 (0.30, 0.55) |
| 2011 | Prior | 177 | 570 | 1.72 (1.48, 1.99) | 1.06 (1.05, 1.07) | 2.09 (1.76, 2.48) |
| 2011 | Study | 111 | 773 | 1.82 (1.59, 2.08) | 1.08 (1.07, 1.08) | 1.00 (0.82, 1.22) |
| 2011 | PERR | | | | | 0.48 (0.39, 0.60) |

Table 17: *Results for the number of admissions for myocardial infarction (MI); and the prior, study period and PERR-adjusted hazard ratios (95% CIs) of gender, age and influenza vaccination group for each annual cohort*

#### *6.4.2.2 Weighted results and PERR adjustment*

From the IPTW analysis, adjusting for measured confounders, the results were variously distributed around unity (Table 18). No results for any cohort's study period (blue circles in Figure 29) were significantly different from the null, but the greatest protective effect was seen in 2003 with an HR of 0.78 (95% CI: 0.59, 1.04), respectively. From 2004 onwards, all HRs were above unity indicating a harmful effect with HRs as high as 1.28 (95% CI: 0.95, 1.74) in 2009. Once the PERR method had been applied, the weighted results (blue triangles in Figure 29) were more commensurate with those from the PERR-adjusted unweighted-Cox models (grew triangles in Figure 29), including the non-significant effect estimated for the 2001 cohort. The diagnostic plots in 6.6 Appendix A of the standardised and unstandardised mean differences of potential confounders in the study periods revealed better balanced was generally achieved for the earlier cohorts, noting that not all the variables contributed to the propensity score. It was also apparent that after 2001, age became the leading variable with the greatest imbalance between vaccination groups. This seemed to coincide with the shift from risk-based vaccination prior to the policy introduction to the age-based eligibility criterion for vaccination.

| Cohort | HR for each period | | PERR HR |
|---|---|---|---|
| | Prior | Study | |
| 1997 | 1.63 (1.17, 2.28) | 1.00 (0.65, 1.55) | 0.61 (0.36, 1.03) |
| 1998 | 1.86 (1.41, 2.44) | 0.88 (0.62, 1.25) | 0.48 (0.31, 0.73) |
| 1999 | 2.25 (1.80, 2.81) | 0.99 (0.79, 1.24) | 0.44 (0.32, 0.58) |
| 2000 | 1.22 (1.02, 1.47) | 0.86 (0.72, 1.01) | 0.70 (0.55, 0.91) |
| 2001 | 1.36 (1.10, 1.67) | 1.09 (0.88, 1.34) | 0.80 (0.60, 1.07) |
| 2002 | 1.61 (1.25, 2.08) | 1.04 (0.81, 1.33) | 0.64 (0.47, 0.88) |
| 2003 | 1.79 (1.42, 2.27) | 0.78 (0.59, 1.04) | 0.44 (0.29, 0.61) |
| 2004 | 1.77 (1.42, 2.22) | 1.22 (0.94, 1.59) | 0.69 (0.48, 0.93) |
| 2005 | 1.66 (1.34, 2.04) | 1.08 (0.86, 1.35) | 0.65 (0.48, 0.90) |
| 2006 | 2.07 (1.62, 2.64) | 1.17 (0.85, 1.61) | 0.57 (0.39, 0.79) |
| 2007 | 2.55 (1.92, 3.38) | 1.11 (0.81, 1.52) | 0.44 (0.29, 0.66) |
| 2008 | 1.61 (1.24, 2.09) | 1.07 (0.77, 1.50) | 0.67 (0.44, 0.99) |
| 2009 | 2.05 (1.61, 2.59) | 1.28 (0.95, 1.74) | 0.63 (0.42, 0.89) |
| 2010 | 2.76 (2.14, 3.55) | 1.24 (0.88, 1.75) | 0.45 (0.29, 0.65) |
| 2011 | 2.09 (1.70, 2.58) | 1.00 (0.78, 1.28) | 0.48 (0.34, 0.64) |

*Table 18:* Inverse probability treatment weighted hazard ratios (95% CIs) for the study and prior periods of each cohort from 1997 to 2011. The PERR results are those from the adjustment of the weighted HR for the study adjusted with that of the prior periods, presented with 95% booststrapped confidence intervals.

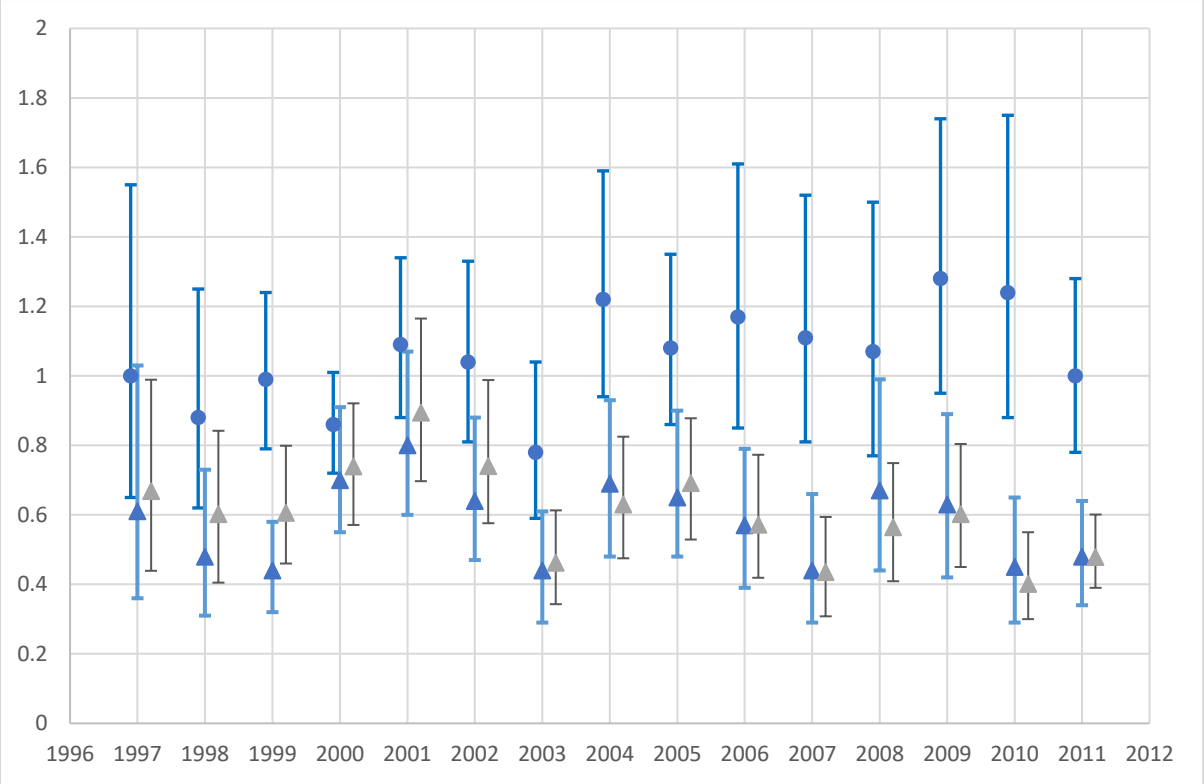


*Figure 29:* Plots for the cohorts from years 1997 to 2011, comparing the inverse probability treatment weighted hazard ratios from the study periods (blue circles) of each, from the PERR adjustment for the weighted HRs (blue triangles) and from the PERR results of the unweighted survival analyses (grey triangles).

### *6.4.2.3 PERR adjustment of aggregated results*

Having aggregated the data from across the annual cohorts, the PERR-adjusted HR of 0.61 (95% CI: 0.56, 0.66) for vaccination against MIs from the global model indicated that the average reduction in risk from 1997 to 2011 was 39% (95% CI: 34%, 44%) (Table 19). In the global model, the confounding effect of age and gender was seen to be more stable in the prior and study periods, with the point estimates for the HR for age being around 1.05 to 1.06 and that of gender ranging only from 1.60 in the prior to 1.67 in the study period.

| Period | No. of MI | HR gender | HR age | HR FV group |
|---|---|---|---|---|
| Prior | 7828 | 1.60 (1.52, 1.69) | 1.05 (1.05, 1.05) | 1.93 (1.83, 2.04) |
| Study | 9871 | 1.67 (1.59, 1.75) | 1.06 (1.06, 1.07) | 1.18 (1.12, 1.25) |
| PERR | | | | 0.61 (0.56, 0.66) |

*Table 19:* Number of myocardial infarctions (MI) and hazard ratios (95% confidence intervals) for MI by influenza vaccination group from the global model aggregating all annual cohorts, allowing for clustering around patient

### *6.4.2.4 Pairwise estimates*

The HRs of vaccination effect from the pairwise model, in which age, gender and period were adjusted for, were all further from unity than the PERR estimates (HR of 0.38 in 2003) (Table 20; Figure 28). This would indicate a marginally greater overall protective effect of vaccination against MI than that estimated through the PERR method, although the overlap of confidence intervals would suggest this to be non-significant.

| Cohort | PERR-TVC | Pairwise cohort | Pairwise HR |
|---|---|---|---|
| 1997 | 0.55 (0.30, 1.07) | 656 | 0.51 (0.33, 0.79) |
| 1998 | 0.59 (0.33, 0.99) | 888 | 0.56 (0.38, 0.83) |
| 1999 | 0.54 (0.34, 0.78) | 960 | 0.55 (0.40, 0.76) |
| 2000 | 0.78 (0.58, 1.08) | 1053 | 0.64 (0.49, 0.83) |
| 2001 | 0.80 (0.56, 1.13) | 894 | 0.80 (0.59, 1.08) |
| 2002 | 0.64 (0.38, 1.03) | 797 | 0.70 (0.49, 1.01) |
| 2003 | 0.48 (0.27, 0.75) | 874 | 0.38 (0.27, 0.54) |
| 2004 | 0.62 (0.41, 1.05) | 933 | 0.53 (0.38, 0.72) |
| 2005 | 0.67 (0.47, 0.99) | 1020 | 0.53 (0.40, 0.71) |
| 2006 | 0.66 (0.38, 1.04) | 938 | 0.49 (0.34, 0.70) |
| 2007 | 0.43 (0.25, 0.67) | 995 | 0.39 (0.27, 0.56) |
| 2008 | 0.63 (0.40, 0.91) | 1137 | 0.46 (0.33, 0.64) |
| 2009 | 0.65 (0.42, 0.98) | 1241 | 0.53 (0.39, 0.72) |
| 2010 | 0.41 (0.26, 0.63) | 1278 | 0.38 (0.27, 0.52) |
| 2011 | 0.48 (0.32, 0.71) | 1566 | 0.44 (0.33, 0.58) |

*Table 20:* Hazard ratios (bootstrapped 95% CIs) for each annual cohort for the effect of influenza vaccination on MI hospital admissions from the pairwise model and the PERR-adjusted TVC model. As the data are analysed using the pairwise method for a subset of patients from each annual cohort with outcomes in either the prior or study period, then the size of this subset is given in the table as “Pairwise N”. The size of the PERR-TVC cohort remains the same as that of the standard PERR-adjusted models.

### *6.4.2.5 Time-varying covariate models*

The estimates from the PERR-adjusted TVC model broadly followed the same trend as those from the PERR-adjusted Cox and the Pairwise models, and were closer to the null than the Pairwise results in all, but four of the years (Table 20; Figure 28). However, there was greater imprecision around the estimates, with confidence intervals that were wider than those from the PERR and Pairwise methods.

## 6.4.3 Effectiveness of vaccination on influenza

The HRs for age in the Cox model of every cohort's study period were significantly and consistently greater than unity, indicating an increasing risk of hospitalisation or treatment for influenza with age (Table 21). Conversely, the HR for males was significantly greater than unity up to 2002, after which it decreased to below one, but not significantly so - indicating at least parity in the risks between the genders. The HR for vaccination status was significantly in excess of one in every cohort, ranging between 1.31 in 2005 and 1.82 in 2011. The HRs in the vaccine-free prior periods were greater than the study periods, indicating pre-existing bias, except for the 2001 cohort, which with an HR of 1.48 was only 0.02 greater than that of the study period (

| Cohort | Period | No. events | Hospital admissions as % of events | Hazard ratios | | |
|---|---|---|---|---|---|---|
| | | | | Gender | Age | FV group |
| 1997 | Prior | 2747 | 5.8% | 1.02 (0.94, 1.10) | 1.02 (1.01, 1.02) | 2.01 (1.84, 2.20) |
| 1997 | Study | 2548 | 15.0% | 1.12 (1.03, 1.21) | 1.03 (1.02, 1.03) | 1.53 (1.38, 1.69) |
| 1997 | PERR | | | | | 0.76 (0.67, 0.86) |
| 1998 | Prior | 2692 | 13.2% | 1.06 (0.98, 1.14) | 1.02 (1.02, 1.03) | 2.07 (1.87, 2.29) |
| 1998 | Study | 2627 | 23.9% | 1.17 (1.08, 1.26) | 1.04 (1.03, 1.04) | 1.46 (1.30, 1.64) |
| 1998 | PERR | | | | | 0.71 (0.61, 0.82) |
| 1999 | Prior | 2748 | 17.0% | 1.16 (1.07, 1.25) | 1.03 (1.02, 1.03) | 2.12 (1.94, 2.32) |
| 1999 | Study | 2780 | 27.4% | 1.10 (1.02, 1.19) | 1.04 (1.03, 1.04) | 1.48 (1.34, 1.63) |
| 1999 | PERR | | | | | 0.70 (0.62, 0.79) |
| 2000 | Prior | 2741 | 17.7% | 1.06 (0.98, 1.15) | 1.03 (1.03, 1.04) | 1.65 (1.53, 1.78) |
| 2000 | Study | 2621 | 22.8% | 1.12 (1.03, 1.21) | 1.04 (1.03, 1.04) | 1.37 (1.27, 1.48) |
| 2000 | PERR | | | | | 0.83 (0.74, 0.93) |
| 2001 | Prior | 1831 | 16.5% | 1.01 (0.92, 1.11) | 1.03 (1.02, 1.04) | 1.48 (1.34, 1.63) |
| 2001 | Study | 2000 | 29.2% | 1.12 (1.02, 1.22) | 1.04 (1.04, 1.05) | 1.46 (1.33, 1.61) |
| 2001 | PERR | | | | | 0.99 (0.87, 1.12) |
| 2002 | Prior | 1722 | 17.8% | 1.02 (0.93, 1.13) | 1.03 (1.03, 1.04) | 1.87 (1.68, 2.08) |
| 2002 | Study | 1759 | 33.1% | 1.02 (0.92, 1.12) | 1.04 (1.04, 1.05) | 1.44 (1.28, 1.61) |
| 2002 | PERR | | | | | 0.77 (0.67, 0.90) |
| 2003 | Prior | 1856 | 19.7% | 0.88 (0.80, 0.97) | 1.03 (1.03, 1.04) | 2.00 (1.81, 2.22) |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2003 | Study | 1911 | 32.9% | 0.98 (0.89, 1.07) | 1.04 (1.04, 1.05) | 1.46 (1.31, 1.63) |
| 2003 | PERR | | | | | 0.73 (0.64, 0.84) |
| 2004 | Prior | 2187 | 20.8% | 0.90 (0.83, 0.98) | 1.04 (1.03, 1.04) | 1.97 (1.79, 2.16) |
| 2004 | Study | 2420 | 32.0% | 1.03 (0.94, 1.12) | 1.04 (1.04, 1.05) | 1.53 (1.39, 1.69) |
| 2004 | PERR | | | | | 0.78 (0.70, 0.89) |
| 2005 | Prior | 2735 | 18.5% | 0.94 (0.87, 1.01) | 1.03 (1.03, 1.04) | 1.70 (1.57, 1.84) |
| 2005 | Study | 2697 | 32.3% | 1.03 (0.96, 1.12) | 1.04 (1.04, 1.05) | 1.31 (1.20, 1.43) |
| 2005 | PERR | | | | | 0.77 (0.69, 0.86) |
| 2006 | Prior | 2553 | 20.6% | 0.98 (0.91, 1.07) | 1.04 (1.03, 1.04) | 2.13 (1.94, 2.34) |
| 2006 | Study | 2597 | 31.6% | 0.99 (0.91, 1.07) | 1.04 (1.04, 1.05) | 1.58 (1.42, 1.74) |
| 2006 | PERR | | | | | 0.74 (0.66, 0.83) |
| 2007 | Prior | 2867 | 18.5% | 0.95 (0.88, 1.02) | 1.03 (1.02, 1.03) | 1.76 (1.60, 1.93) |
| 2007 | Study | 2793 | 32.3% | 1.02 (0.94, 1.10) | 1.04 (1.04, 1.05) | 1.47 (1.33, 1.64) |
| 2007 | PERR | | | | | 0.84 (0.73, 0.96) |
| 2008 | Prior | 3280 | 18.8% | 0.92 (0.86, 0.99) | 1.03 (1.03, 1.04) | 1.80 (1.66, 1.96) |
| 2008 | Study | 3291 | 35.7% | 0.96 (0.89, 1.03) | 1.04 (1.04, 1.05) | 1.35 (1.23, 1.47) |
| 2008 | PERR | | | | | 0.75 (0.66, 0.83) |
| 2009 | Prior | 3586 | 21.5% | 0.94 (0.88, 1.01) | 1.03 (1.03, 1.03) | 1.92 (1.77, 2.08) |
| 2009 | Study | 3144 | 34.8% | 0.97 (0.90, 1.04) | 1.04 (1.04, 1.05) | 1.47 (1.34, 1.62) |
| 2009 | PERR | | | | | 0.77 (0.69, 0.86) |
| 2010 | Prior | 3561 | 23.9% | 0.91 (0.85, 0.97) | 1.03 (1.03, 1.04) | 1.92 (1.77, 2.09) |
| 2010 | Study | 3490 | 37.5% | 0.97 (0.90, 1.03) | 1.04 (1.04, 1.05) | 1.80 (1.66, 1.97) |
| 2010 | PERR | | | | | 0.94 (0.85, 1.05) |
| 2011 | Prior | 4025 | 24.3% | 0.95 (0.89, 1.02) | 1.04 (1.03, 1.04) | 2.24 (2.09, 2.41) |
| 2011 | Study | 4127 | 48.6% | 0.98 (0.92, 1.04) | 1.05 (1.05, 1.05) | 1.82 (1.69, 1.96) |
| 2011 | PERR | | | | | 0.81 (0.73, 0.89) |

Table 21; Figure 30). The PERR-adjusted HRs ranged from 0.70 in 1999 reaching the maximum of 0.99 in 2001. The HRs from the pairwise method, while always further below the null, tracked very closely with the PERR-adjusted results, and ranged from 0.63 in 1999 to 0.90 in 2001 (Table 22; Figure 31). While the prevalence of influenza outcomes remained relatively stable varying between 3 and 4% in each cohort, the number of hospital admissions as a proportion of the composite influenza was 15% in 1997 and after a temporary fall in 2000, eventually rose to 48.6% in 2011 (Table 21). While this increase was quite marked, it seemed to bear no relation to the trends in the effect of vaccination, either before or after PERR adjustment (Figure 31).

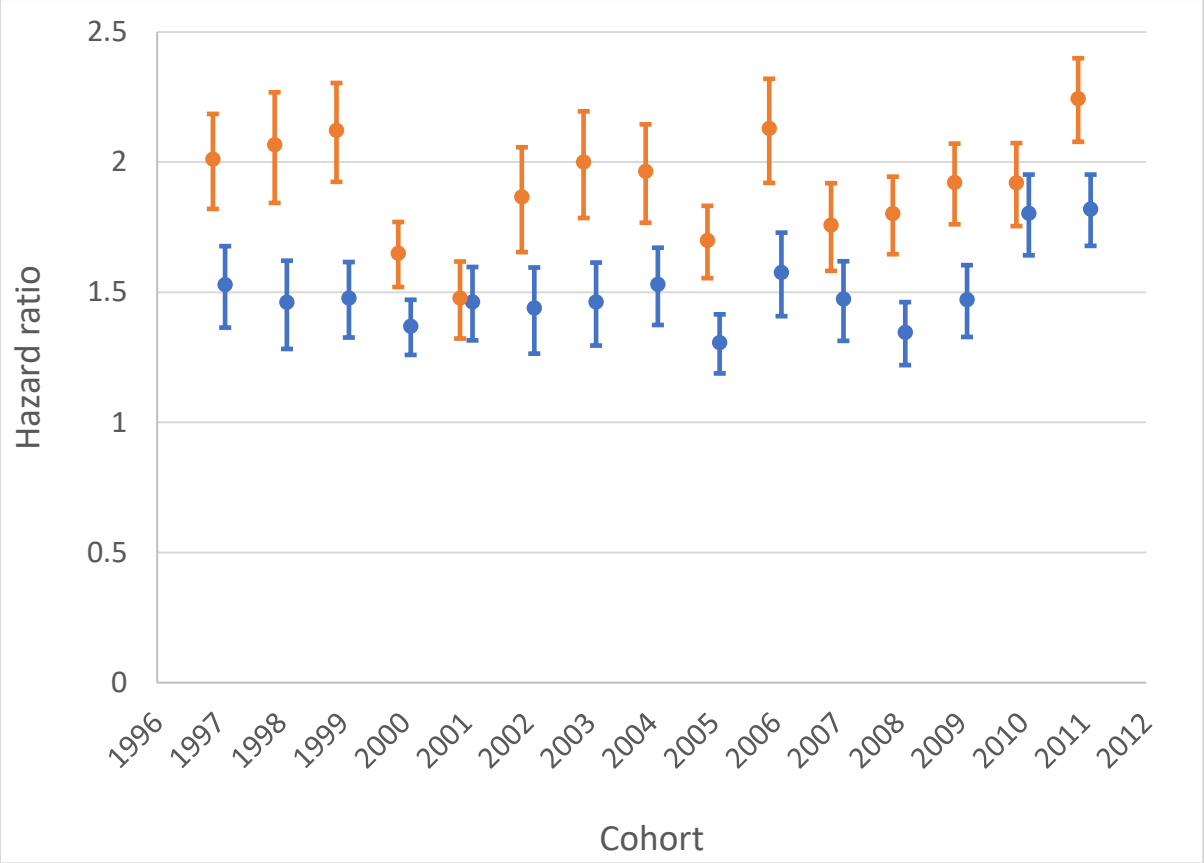


*Figure 30:* Plot of hazard ratios of the effect on the composite influenza outcome of influenza vaccination from the Cox regression models adjusted for age and gender for the prior (orange) and study (blue) periods of each annual cohort.

| Cohort | Period | No. events | Hospital admissions as % of events | Hazard ratios | | |
|---|---|---|---|---|---|---|
| | | | | Gender | Age | FV group |
| 1997 | Prior | 2747 | 5.8% | 1.02 (0.94, 1.10) | 1.02 (1.01, 1.02) | 2.01 (1.84, 2.20) |
| 1997 | Study | 2548 | 15.0% | 1.12 (1.03, 1.21) | 1.03 (1.02, 1.03) | 1.53 (1.38, 1.69) |
| 1997 | PERR | | | | | 0.76 (0.67, 0.86) |
| 1998 | Prior | 2692 | 13.2% | 1.06 (0.98, 1.14) | 1.02 (1.02, 1.03) | 2.07 (1.87, 2.29) |
| 1998 | Study | 2627 | 23.9% | 1.17 (1.08, 1.26) | 1.04 (1.03, 1.04) | 1.46 (1.30, 1.64) |
| 1998 | PERR | | | | | 0.71 (0.61, 0.82) |
| 1999 | Prior | 2748 | 17.0% | 1.16 (1.07, 1.25) | 1.03 (1.02, 1.03) | 2.12 (1.94, 2.32) |
| 1999 | Study | 2780 | 27.4% | 1.10 (1.02, 1.19) | 1.04 (1.03, 1.04) | 1.48 (1.34, 1.63) |
| 1999 | PERR | | | | | 0.70 (0.62, 0.79) |
| 2000 | Prior | 2741 | 17.7% | 1.06 (0.98, 1.15) | 1.03 (1.03, 1.04) | 1.65 (1.53, 1.78) |
| 2000 | Study | 2621 | 22.8% | 1.12 (1.03, 1.21) | 1.04 (1.03, 1.04) | 1.37 (1.27, 1.48) |
| 2000 | PERR | | | | | 0.83 (0.74, 0.93) |
| 2001 | Prior | 1831 | 16.5% | 1.01 (0.92, 1.11) | 1.03 (1.02, 1.04) | 1.48 (1.34, 1.63) |
| 2001 | Study | 2000 | 29.2% | 1.12 (1.02, 1.22) | 1.04 (1.04, 1.05) | 1.46 (1.33, 1.61) |
| 2001 | PERR | | | | | 0.99 (0.87, 1.12) |
| 2002 | Prior | 1722 | 17.8% | 1.02 (0.93, 1.13) | 1.03 (1.03, 1.04) | 1.87 (1.68, 2.08) |
| 2002 | Study | 1759 | 33.1% | 1.02 (0.92, 1.12) | 1.04 (1.04, 1.05) | 1.44 (1.28, 1.61) |
| 2002 | PERR | | | | | 0.77 (0.67, 0.90) |
| 2003 | Prior | 1856 | 19.7% | 0.88 (0.80, 0.97) | 1.03 (1.03, 1.04) | 2.00 (1.81, 2.22) |
| 2003 | Study | 1911 | 32.9% | 0.98 (0.89, 1.07) | 1.04 (1.04, 1.05) | 1.46 (1.31, 1.63) |
| 2003 | PERR | | | | | 0.73 (0.64, 0.84) |
| 2004 | Prior | 2187 | 20.8% | 0.90 (0.83, 0.98) | 1.04 (1.03, 1.04) | 1.97 (1.79, 2.16) |
| 2004 | Study | 2420 | 32.0% | 1.03 (0.94, 1.12) | 1.04 (1.04, 1.05) | 1.53 (1.39, 1.69) |
| 2004 | PERR | | | | | 0.78 (0.70, 0.89) |
| 2005 | Prior | 2735 | 18.5% | 0.94 (0.87, 1.01) | 1.03 (1.03, 1.04) | 1.70 (1.57, 1.84) |
| 2005 | Study | 2697 | 32.3% | 1.03 (0.96, 1.12) | 1.04 (1.04, 1.05) | 1.31 (1.20, 1.43) |
| 2005 | PERR | | | | | 0.77 (0.69, 0.86) |
| 2006 | Prior | 2553 | 20.6% | 0.98 (0.91, 1.07) | 1.04 (1.03, 1.04) | 2.13 (1.94, 2.34) |
| 2006 | Study | 2597 | 31.6% | 0.99 (0.91, 1.07) | 1.04 (1.04, 1.05) | 1.58 (1.42, 1.74) |
| 2006 | PERR | | | | | 0.74 (0.66, 0.83) |
| 2007 | Prior | 2867 | 18.5% | 0.95 (0.88, 1.02) | 1.03 (1.02, 1.03) | 1.76 (1.60, 1.93) |
| 2007 | Study | 2793 | 32.3% | 1.02 (0.94, 1.10) | 1.04 (1.04, 1.05) | 1.47 (1.33, 1.64) |
| 2007 | PERR | | | | | 0.84 (0.73, 0.96) |
| 2008 | Prior | 3280 | 18.8% | 0.92 (0.86, 0.99) | 1.03 (1.03, 1.04) | 1.80 (1.66, 1.96) |
| 2008 | Study | 3291 | 35.7% | 0.96 (0.89, 1.03) | 1.04 (1.04, 1.05) | 1.35 (1.23, 1.47) |
| 2008 | PERR | | | | | 0.75 (0.66, 0.83) |
| 2009 | Prior | 3586 | 21.5% | 0.94 (0.88, 1.01) | 1.03 (1.03, 1.03) | 1.92 (1.77, 2.08) |
| 2009 | Study | 3144 | 34.8% | 0.97 (0.90, 1.04) | 1.04 (1.04, 1.05) | 1.47 (1.34, 1.62) |
| 2009 | PERR | | | | | 0.77 (0.69, 0.86) |
| 2010 | Prior | 3561 | 23.9% | 0.91 (0.85, 0.97) | 1.03 (1.03, 1.04) | 1.92 (1.77, 2.09) |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2010 | Study | 3490 | 37.5% | 0.97 (0.90, 1.03) | 1.04 (1.04, 1.05) | 1.80 (1.66, 1.97) |
| 2010 | PERR | | | | | 0.94 (0.85, 1.05) |
| 2011 | Prior | 4025 | 24.3% | 0.95 (0.89, 1.02) | 1.04 (1.03, 1.04) | 2.24 (2.09, 2.41) |
| 2011 | Study | 4127 | 48.6% | 0.98 (0.92, 1.04) | 1.05 (1.05, 1.05) | 1.82 (1.69, 1.96) |
| 2011 | PERR | | | | | 0.81 (0.73, 0.89) |

*Table 21:* Results for the number of (composite influenza) outcomes; percentage of outcomes that were hospital admissions for suspected influenza; and the prior, study period and PERR-adjusted hazard ratios (95% CIs) of gender, age and influenza vaccination group for each annual cohort.

| Cohort | N | % vaccinated | Pairwise HR vaccination |
|---|---|---|---|
| 1997 | 4859 | 19.5% | 0.73 (0.63, 0.84) |
| 1998 | 4869 | 14.4% | 0.65 (0.55, 0.77) |
| 1999 | 5096 | 19.5% | 0.63 (0.54, 0.72) |
| 2000 | 4920 | 48.3% | 0.76 (0.68, 0.86) |
| 2001 | 3540 | 29.7% | 0.90 (0.78, 1.05) |
| 2002 | 3200 | 23.7% | 0.71 (0.60, 0.84) |
| 2003 | 3455 | 25.2% | 0.64 (0.55, 0.76) |
| 2004 | 4227 | 23.9% | 0.72 (0.62, 0.84) |
| 2005 | 5003 | 29.9% | 0.72 (0.63, 0.81) |
| 2006 | 4732 | 20.2% | 0.70 (0.60, 0.81) |
| 2007 | 5216 | 17.2% | 0.79 (0.68, 0.92) |
| 2008 | 6023 | 21.5% | 0.69 (0.61, 0.79) |
| 2009 | 6193 | 19.9% | 0.72 (0.63, 0.82) |
| 2010 | 6486 | 19.3% | 0.86 (0.75, 0.98) |
| 2011 | 7537 | 23.8% | 0.76 (0.68, 0.85) |

*Table 22:* Results (hazard ratios (95% CIs)) from the pairwise analysis of the effect of influenza vaccination on the composite influenza outcome for each cohort. Note that each cohort is a reduced subset of patients with an outcome in either the prior or study period as demanded by the pairwise likelihood.

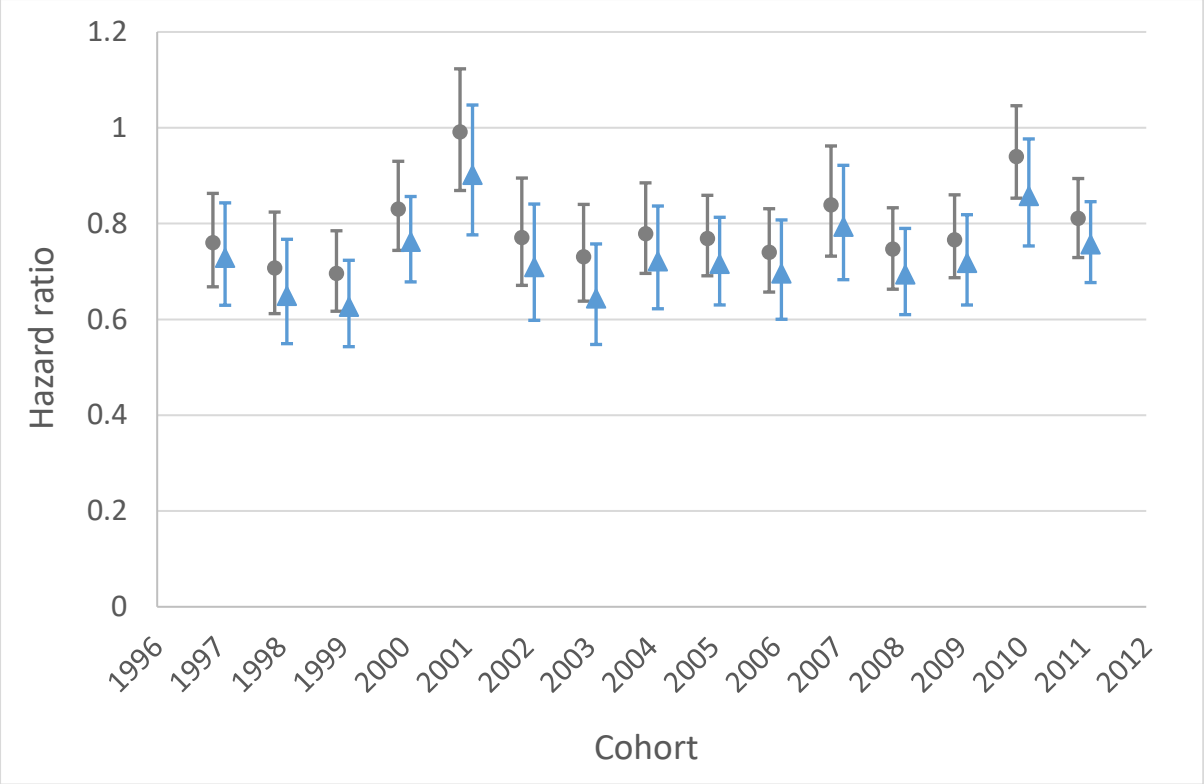


*Figure 31:* Plot of hazard ratios for the estimated effect of influenza vaccination on influenza outcomes from the PERR-adjusted model (grey dots) and the Pairwise model (blue triangles) with errors bars representing the bootstrapped 95% confidence intervals.

### 6.4.4 Sub-group analysis

No significant effect on the influenza outcome was detected from the interaction between age and vaccination status with all the 95% CIs for all HRs including unity (not shown). For the MI outcome, however, there was a significant effect of interaction between vaccination and age in years 2004 and 2008 with HRs above unity and marginal significance in 2010 and 2011 (Table 23 ). All other interactions were above unity, but not significantly so, apart from the 1997, 1998 and 2005 cohorts, for which the main effects of vaccination were also weakest out of all the cohorts. Given that all the point estimates for the vaccination main effect were below one, then the interpretation of the interactions for the 13 cohorts with interaction HRs above one is that for those years, the effectiveness of vaccination appeared to wane with age. For example, the hazard of an MI after vaccination in 2009 was estimated to be 0.55 at 65y, but according to the interaction this had increased to 0.67 at 85y.

| Cohort | HR vaccination | HR vaccine*age |
|---|---|---|
| 1997 | 0.85 (0.23, 1.48) | 0.98 (0.92, 1.04) |
| 1998 | 0.71 (0.18, 1.24) | 0.99 (0.93, 1.04) |
| 1999 | 0.44 (0.20, 0.68) | 1.03 (0.99, 1.08) |
| 2000 | 0.60 (0.39, 0.80) | 1.02 (0.99, 1.05) |
| 2001 | 0.80 (0.46, 1.13) | 1.01 (0.98, 1.04) |
| 2002 | 0.62 (0.33, 0.92) | 1.02 (0.99, 1.05) |
| 2003 | 0.40 (0.22, 0.58) | 1.01 (0.98, 1.05) |
| 2004 | 0.43 (0.25, 0.62) | 1.04 (1.01, 1.07) |
| 2005 | 0.85 (0.51, 1.20) | 0.98 (0.95, 1.01) |
| 2006 | 0.46 (0.24, 0.68) | 1.02 (0.99, 1.06) |
| 2007 | 0.35 (0.19, 0.50) | 1.03 (0.99, 1.06) |
| 2008 | 0.42 (0.24, 0.59) | 1.04 (1.00, 1.07) |
| 2009 | 0.55 (0.35, 0.76) | 1.01 (0.98, 1.04) |
| 2010 | 0.29 (0.17, 0.40) | 1.03 (1.00, 1.06) |
| 2011 | 0.37 (0.24, 0.51) | 1.03 (1.00, 1.05) |

*Table 23:* PERR-adjusted hazard ratios (95% CIs) of vaccination main effect and its interaction with age on myocardial infarctions for annual cohorts, from the model including age and gender main effects and interaction.

In the analysis of influenza vaccine effectiveness by PPV subgroup, a high degree of uncertainty characterised the results from 1997 to 2002 with relatively wide confidence intervals for those years and the point estimates for the HRs for influenza-vaccine status were greater than unity among the PPV recipients in 1998, 2000 and 2001 (Table 24; Figure 32). In 2001, the PERR-adjusted hazard of an influenza outcome following influenza vaccination was significantly greater among the PPV recipients, and greater than unity, indicating evidence of a harmful effect of influenza vaccination in the PPV subgroup. The pairwise-estimated HR for PPV recipients in 2001, however, while greater than the subgroup without PPV, was not significantly greater than unity (Table 24; Figure 33). In contrast the variability in the HRs for influenza vaccination among patients with no record of PPV, whilst fluctuating and of variable significance, were more stable and consistently below one, indicating a protective effect of the influenza vaccine in this sub-group.

In terms of the confounding-adjustment method, there were no appreciable differences in the results after 2001, with consistent overlap thereafter between the 95% confidence intervals of the PERR and Pairwise estimates for vaccination effect. According to both methods, the Pairwise HRs were lower than the PERR-adjusted analogues after 2000 in both PPV subgroups. However, before 2001 the small sample sizes likely led to the large variability and lack of stability in the estimates from both methods for the PPV subgroup.

| Cohort | PPV subgroup | N | % vaccinated | PERR-adjusted HR vaccination | Pairwise-adjusted HR vaccination |
|---|---|---|---|---|---|
| 1997 | PPV | 432 | 32.2% | 0.78 (0.28, 1.88) | 0.67 (0.24, 1.91) |
| 1998 | PPV | 1065 | 26.2% | 1.39 (0.67, 2.67) | 1.44 (0.67, 3.07) |
| 1999 | PPV | 1905 | 37.4% | 0.70 (0.43, 1.10) | 0.57 (0.32, 1.01) |
| 2000 | PPV | 2208 | 60.3% | 1.12 (0.72, 1.77) | 1.09 (0.66, 1.81) |
| 2001 | PPV | 2054 | 58.8% | 1.95 (1.20, 3.27) | 1.77 (0.97, 3.26) |
| 2002 | PPV | 1234 | 31.2% | 0.88 (0.46, 1.55) | 0.79 (0.37, 1.68) |
| 2003 | PPV | 1349 | 31.7% | 0.60 (0.31, 1.07) | 0.56 (0.28, 1.14) |
| 2004 | PPV | 2633 | 30.0% | 0.83 (0.57, 1.26) | 0.92 (0.57, 1.50) |
| 2005 | PPV | 7882 | 44.4% | 0.73 (0.55, 0.96) | 0.68 (0.48, 0.95) |
| 2006 | PPV | 9265 | 23.6% | 0.85 (0.63, 1.16) | 0.81 (0.56, 1.16) |
| 2007 | PPV | 10984 | 19.1% | 1.01 (0.74, 1.37) | 0.90 (0.64, 1.27) |
| 2008 | PPV | 15653 | 24.4% | 0.72 (0.58, 0.90) | 0.62 (0.47, 0.81) |
| 2009 | PPV | 17154 | 20.5% | 0.93 (0.74, 1.17) | 0.96 (0.73, 1.26) |
| 2010 | PPV | 19832 | 17.1% | 1.11 (0.89, 1.39) | 0.98 (0.73, 1.33) |
| 2011 | PPV | 14587 | 21.3% | 0.93 (0.72, 1.21) | 0.88 (0.66, 1.18) |
| 1997 | no PPV | 60009 | 10.7% | 0.77 (0.68, 0.90) | 0.73 (0.62, 0.85) |
| 1998 | no PPV | 64466 | 6.9% | 0.68 (0.58, 0.81) | 0.63 (0.52, 0.77) |
| 1999 | no PPV | 65944 | 9.4% | 0.69 (0.59, 0.80) | 0.63 (0.53, 0.74) |
| 2000 | no PPV | 65715 | 35.5% | 0.86 (0.76, 0.95) | 0.77 (0.68, 0.88) |
| 2001 | no PPV | 53861 | 19.3% | 0.94 (0.80, 1.10) | 0.81 (0.68, 0.95) |
| 2002 | no PPV | 51031 | 14.8% | 0.81 (0.69, 0.94) | 0.74 (0.61, 0.89) |
| 2003 | no PPV | 49184 | 13.1% | 0.80 (0.67, 0.95) | 0.67 (0.55, 0.81) |
| 2004 | no PPV | 47730 | 7.0% | 0.97 (0.79, 1.20) | 0.77 (0.63, 0.94) |
| 2005 | no PPV | 55233 | 9.0% | 1.01 (0.82, 1.22) | 0.86 (0.70, 1.06) |
| 2006 | no PPV | 56996 | 5.6% | 0.77 (0.63, 0.96) | 0.70 (0.56, 0.88) |
| 2007 | no PPV | 64105 | 5.9% | 0.76 (0.61, 0.95) | 0.73 (0.58, 0.92) |
| 2008 | no PPV | 68699 | 8.1% | 0.83 (0.69, 1.00) | 0.78 (0.64, 0.95) |
| 2009 | no PPV | 72342 | 6.7% | 0.74 (0.61, 0.87) | 0.67 (0.55, 0.83) |
| 2010 | no PPV | 76005 | 6.8% | 0.87 (0.73, 1.07) | 0.86 (0.68, 1.09) |
| 2011 | no PPV | 83377 | 9.0% | 0.80 (0.69, 0.94) | 0.79 (0.67, 0.93) |

*Table 24:* PERR-adjusted and Pairwise-estimated hazard ratios (95% CIs) of the effect of influenza vaccination on influenza outcomes by sub-groups of pneumococcal vaccination (PPV) status for each cohort.

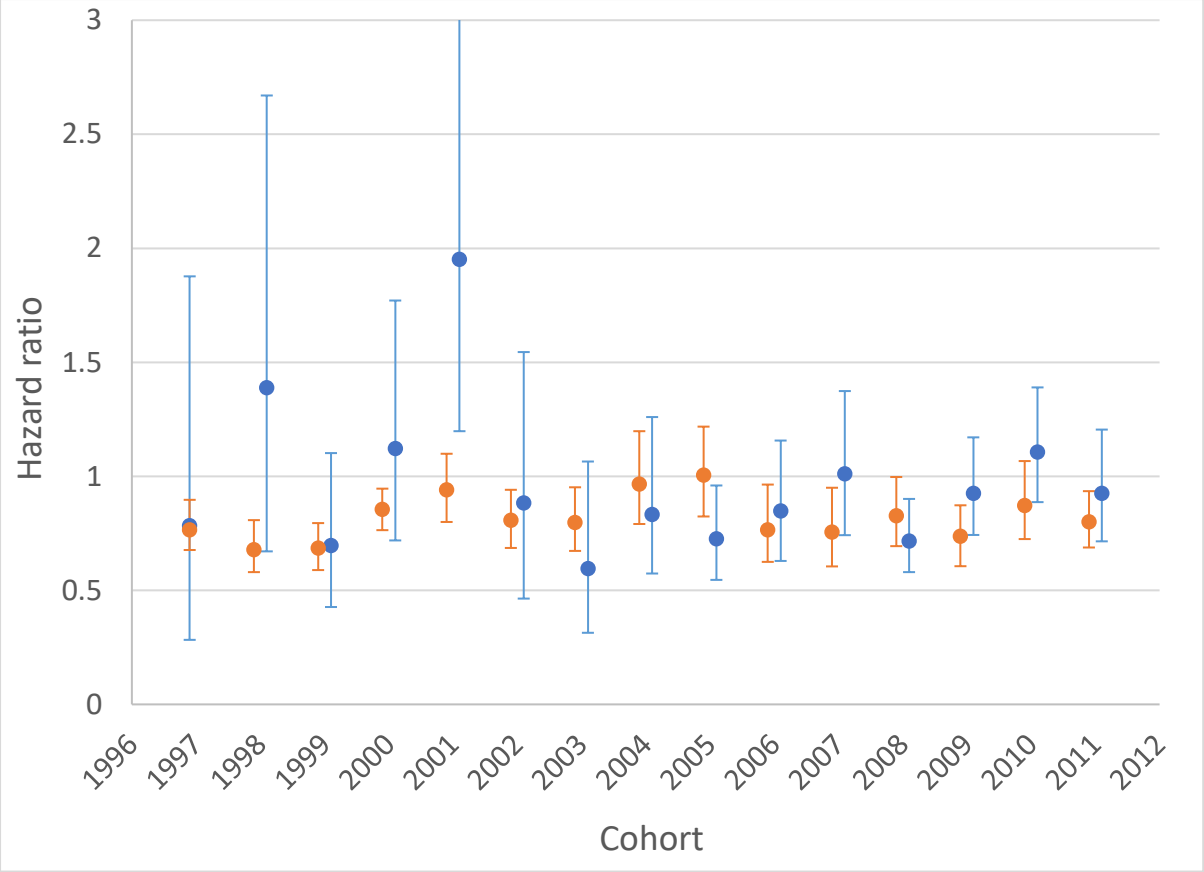


*Figure 32:* Plot of PERR-adjusted hazard ratios of effect of influenza vaccination on influenza outcomes for recipients of the pneumococcal (PPV) vaccination (blue) and patients without PPV (orange).

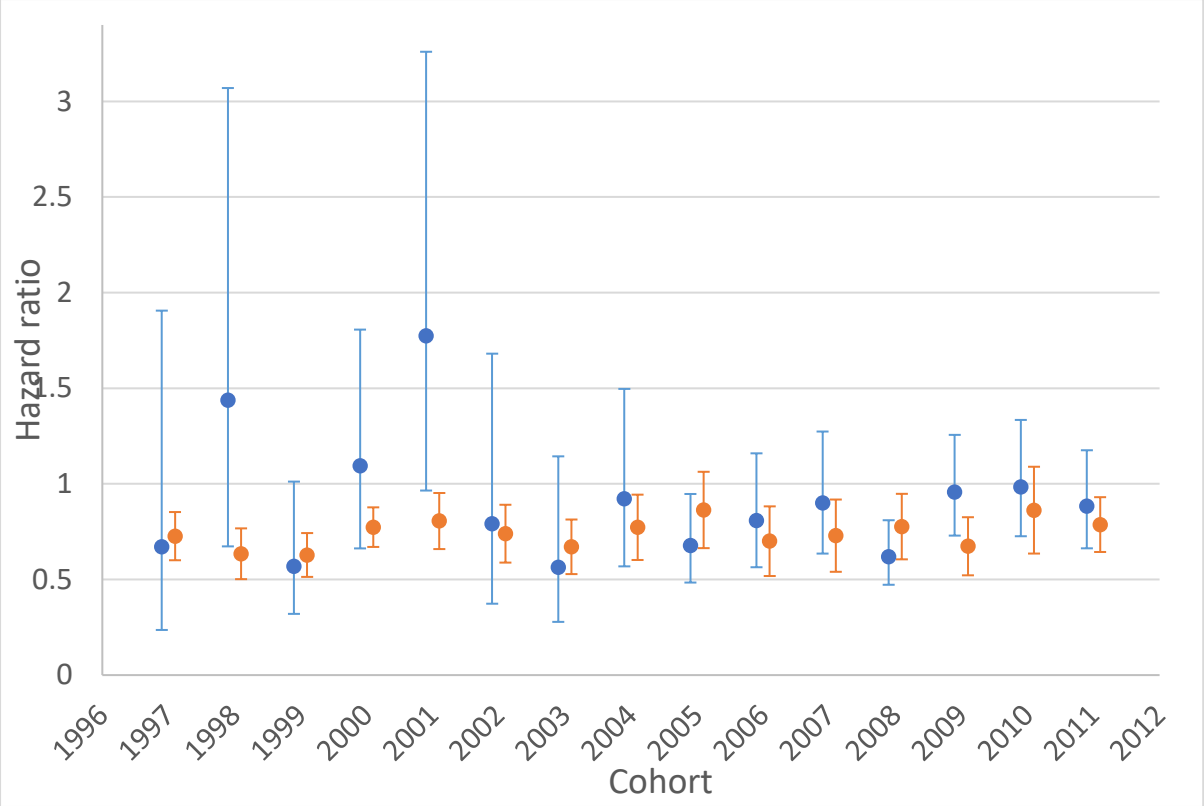


*Figure 33:* Plot of Pairwise-estimated hazard ratios of effect of influenza vaccination on influenza outcomes for recipients of the pneumococcal (PPV) vaccination (blue) and patients without PPV (orange).

## 6.5 Discussion

By accounting for pre-existing confounding biases, the overall effectiveness of the influenza vaccine in reducing the probability of myocardial hospital admissions was estimated on average to be 39% (95% CI: 34, 44) among patients aged at least 65y in UK primary care for years 1997 to 2011. This was reasonably consistent with the estimate of 33% for the effectiveness against MI reported by the FLUCAD randomised study [310] captured in the most recent Cochrane review, although their precision was affected by having fewer patients and therefore fewer outcomes. The results from our study are also broadly in line with the conclusion of a protective effect against major adverse cardiovascular events as estimated in a recent systematic review [312]. Other sources of evidence are available, but using different approaches, in which confounding bias may not have been addressed [303,311]. Aggregating the number of events and the time at risk over all cohorts, the average percentage risk over 365 years in the control group was calculated to be 0.80%. Treating this as the baseline risk, and assuming a relative risk of 0.61 among the vaccinated given by the aggregated PERR results, this corresponded to a risk difference of 0.31% - the small difference a consequence of the low incidence of MI in this selected group of older adults. Although the reduction in the hazard of MI indicated at least moderate effectiveness of the vaccine, expressed as a difference, the vaccine may only be expected to prevent about 3 cases in every 1000 person-years. Having mitigated for unmeasured confounding through the PERR method, the estimated annual vaccine effectiveness varied between 11% in 2001 and 60% in 2010. The results from the Pairwise method followed a broadly similar pattern in the trend over time, but estimated greater effectiveness for all years ranging from 20% in 2001 to 62% in 2003.

The degree to which the hazard ratios from the prior periods deviated from unity was inferred to be an estimate of the size of pre-existing confounding bias between the exposure groups. If there were no confounding bias, and no missing covariates in the Cox model, then the hazard ratios of the prior periods would be all aligned with unity under the assumption of time-invariant confounding. However, these were in excess of unity, suggesting that vaccinated patients in this population newly

vaccinated patients were at a considerably higher risk of both MI and influenza in the prior period.

The results from the IPTW analysis, fluctuated around unity up to 2003, but afterwards, were consistently above unity, indicating an elevated risk of MI as much as 40% and a potentially harmful effect of vaccination. In the context of a one-sided hypothesis, that the vaccination is unlikely to be harmful, this would indicate the presence of residual confounding bias, that had not been resolved through weighing. The presence of bias, however, was more apparent in the weighted estimates for the prior periods, which were all above unity, and significantly so, except for that pertaining to the study in 2000. While these estimates in the prior period were closer to the null than those from the unweighed analysis, the significant difference from the null would suggest considerable bias remained unadjusted by the weighted method. As the modelling was involved in deriving the weights, then the bias could concievably have been further exacerbated by misspeficiation of the propensity-score model, from which the weights were predicted. Once the PERR was applied to the IPTW results, these fell below unity and into close agreement with those from the PERR-adjusted results of the unweighted survival analyses.

Estimates derived through PERR method have been shown to be susceptible to bias from missing covariates [225], and this is in contrast to the within-patient comparison of the Pairwise method. Relative to the Pairwise method, the PERR-adjusted results were consistently closer to the null, but with considerable overlap in their confidence intervals for most years. This may suggest a relatively minor impact from missing covariates and differences in performance, although comparison of the methods without information on all key covariates is complex, and the presence of confounder interactions and strong indication by prior events may have equally led to bias in the Pairwise method.

Instability in the degree of confounding was apparent from the fluctation in the prior estimates over the years of the study. This was also evident in the weighted estimates for the prior periods, which followed a similar trend. This could have been due either to variation in the source of unmeasured confounding, or exogenous fluctuations in the outcomes, due to, say, influenza virulence. If the latter were the case, then this could represent a serious departure from the assumption of a time-

invariant, stable effect of confounding between the prior and study period. A global estimate of vaccine effectiveness was presented as a crude average over year-to-year variability in the confounders and also in vaccine effectiveness. This was derived from using the PERR method to adjust the estimate from the aggregated study period data with the aggregated prior periods, having adjusted simply for age and gender. Robust variance estimates clustered by patients were specified in the Cox models to accommodate a lack of independence between the observations, due to multiple observations from the same patients over different cohorts.

The same between-cohort variability in the prior periods was also manifest in the models of the vaccine effectiveness against influenza outcomes. This, too, indicated instability in the degree and source of confounding bias over time. However, compared to the MI outcomes, the effect of gender on influenza outcomes was less. Effectiveness against influenza varied between 30% in 1999 and no effect in 2001 for the PERR estimates, and between 37% and 10% for the pairwise estimates of the same years, respectively. However, the pattern in effectiveness against influenza exhibited over time broadly agreed with the effectiveness against MIs, including the apparent reduction in the vaccine effectiveness in 2001. If MI is a hypothesised consequence of influenza infection then it would be reasonable to expect to see the same trends in effectiveness against MI and influenza. Since the pattern of the trends in vaccine effectiveness against both outcomes over time were observed to be remarkably similar, this provides further evidence to support the hypothesis that the vaccine may offer a secondary protection against MIs through its mediation of influenza infection.

The reduction in vaccine effectiveness against influenza and MI observed in 2001 was the local minimum of a trend that began in 1998. This coincided with the change from recommending the vaccine according to risk to the introduction of the age-based policy in 1998 and 2000 - a likely cause of the instability in the confounding adjustment for that period. A previous study found no evidence of an effect on excess mortality due to influenza in the age groups that were the subject of the change in policy [313]. However, patient characteristics in the vaccination group would have changed as a result of the policy, which may explain the subsequent reduction of effectiveness in 2001. Certainly, changes in the differential prevalences between vaccine recipients and controls of major cardiovascular diseases were evident in the

data. The greatest differences in hyptertension and coronary heart disease rates were seen in the earlier cohorts, but the prevalences were nearly equalised by the later cohorts. Simultaneously, the differences in age between vaccine recipients and controls grew with each annual cohort. Additionally the diagnostic statistics of the balancing variables used in the weighted analyses revealed that after 2002, age became the most imbalanced variable. These patterns perhaps further reflected the change from the ad-hoc practice of recommending the vaccine based on risk group to the age-based policy. After the period of transition to the age-based policy, the hierarchy of balancing variables, ordered by mean differences, stabilised. Therefore, variability in confounding bias for those years were perhaps more likely determined by antigen evolution and the changes in the mix of influenza viruses, rather than changes in the strength of confounding.

Analysis of effectiveness by PPV sub-groups revealed the influenza vaccine to be protective against influenza in the PPV-free sub-group to varying degrees of statistical significance that were not noticeably dissimilar to influenza-vaccine effectiveness regardless of PPV status. However, there were relatively far fewer patients in the sub group of PPV recipients, especially before 2005. With this in mind, the estimates, which indicated a harmful effect of vaccination (significant for the PERR-adjusted estimate in 2001), could be subject to false-positive, type 1 errors, exacerbated by instability in the confounding adjustments for that period. Furthermore, if such a deleterious drug interaction between vaccines were to exist, it is unlikely that it would be constrained to a few select years.

### 6.5.1 Limitations of this study

The effect of vaccination may partly be determined by the degree of antigenic mismatch between vaccine and virus in any particular season [314]. A weakness of the analysis is that the models did not account for this, although determining the activity of the circulating pathogens and the degree of mismatch between virus and vaccine would likely require further work [315]. This would pose an even greater challenge in operationalising the resulting change in confounding bias, and in how this could be incorporated into the models. In particular, adjustment relies on the conditions for confounding being the same from one year to the next. Using the year before the study to adjust for confounding may be unreasonable, particularly if there was a

large enough change in the type or mix of circulating viruses between the two periods. In addition, determining the true effectiveness against MI may be further complicated [306] by mediation of receptor proteins and differential cardioprotective effect of vaccine strains [316]. Therefore, this presents a potential limitation in applying the before-and-after approach of the PERR to vaccination studies of a seasonal disease such as influenza.

A condition for the validity of the PERR and Pairwise methods is the repeatibility of the outcome. The pathology and treatment for subsequent MIs may be different from the initial MI, which could require a more sophisticated approach to the analysis, stratifying the model for each ordered event, such as in the Prentice-Williams-Peterson model [317]. However, this would not induce repeatability, but it could further encumber adjustment for unmeasured confounding through the PERR method, as the models for each stratum of the ordered events could be different, and therefore, necessarily different between the prior and study periods. Furthermore, although the data are large, there would be a greater chance of model misspecification as there would then be multiple events counted from periods of just one year.

Another potential limitation is the threat to the external validitiy of the cohort from the exclusion criteria that were imposed to strengthen the internal validity of the results: The criterion of selecting only patients with no record of vaccination in the five years preceding the study period was chosen to provide a sufficient washout period for contamination from previous vaccinations, but this may have been too long and may consequently encumbered the generalisability of the results by excluding those patients who had been vaccinated more regularly.

### 6.5.2 Strengths of this study

A strength of this work was the robust approach taken to confounding and the use of real-world data to estimate the effectiveness of the influenza vaccine against MI. The use of such data is ideally suited to discovering secondary benefits to existing treatments and new prophylaxes. Such studies offer an opportunity to observe the effectiveness of an intervention in practice away from the idealised settings of a randomised clinical trial (RCT). Although the tool of randomisation in a properly designed and compliant RCT is widely accepted as the best guarantor against confounding bias, quasi-experimental methods and study designs offer a means of

adjusting for confounders that may not always be fully observed. The application of two QE methods avoided relying on a single set of assumptions, allowing further insights into the data and the nature of the confounding bias. Studying the periods prior to vaccination also offered an evaluation of the size of confounding bias under the assumption of time-invariance. The QE approach was further supplemented with a high-dimensional adjusted analysis, which not only offered further insight into the bias due to unmeasured confounding, but also allowed diagnoses of the degree of imbalance in the measured confounders.

Another strength of the study was the analysis of influenza as an outcome and possible mediator for MI. The concomitant effect against influenza was further evidence of the causal pathway to MI from influenza, the primary target of the vaccine. Previous studies have suggested using outcomes observed in months outside the influenza season to control for confounding [318–320]. While this may help identify the presence of bias, the conditions for confounding before, during and after the influenza season are unlikely to be the same. Furthermore, circulating pathogens outside the influenza season are likely to exist at such low levels as to introduce further variability and imprecision.

A potentially important finding was that in all but three of the years, the results suggested vaccine effectiveness may wane with age against MI in this population of older adults. However, there was no indication of a consistent age effect on vaccination against influenza, which was in accordance with findings from a test-negative study conducted in the same age group [321]. It is important to note the distinction in this regard between this and previous work on the pneumococcal vaccine, in which a net benefit was apparent in older age groups.

### 6.5.3 Conclusions

In every year from 1997 to 2011, evidence was found of a protective effect against MIs, having accounted for the hazard of MIs in the prior period of each cohort. For most of those years, there was some evidence that this effectiveness may decrease with age. The agreement between adjusted effectiveness estimates against influenza and MIs in each year would suggest that prevention of MIs is at least mediated by prevention of influenza infection. By relying on QE adjustment using prior periods, annual changes in pathogen virulence may threaten the validity of the

estimates from some years. However, in aggregate the estimated vaccine effectiveness of 39% was in reasonable concordance with available trial evidence.

## 6.6 Appendix A – IPTW diagnostic plots

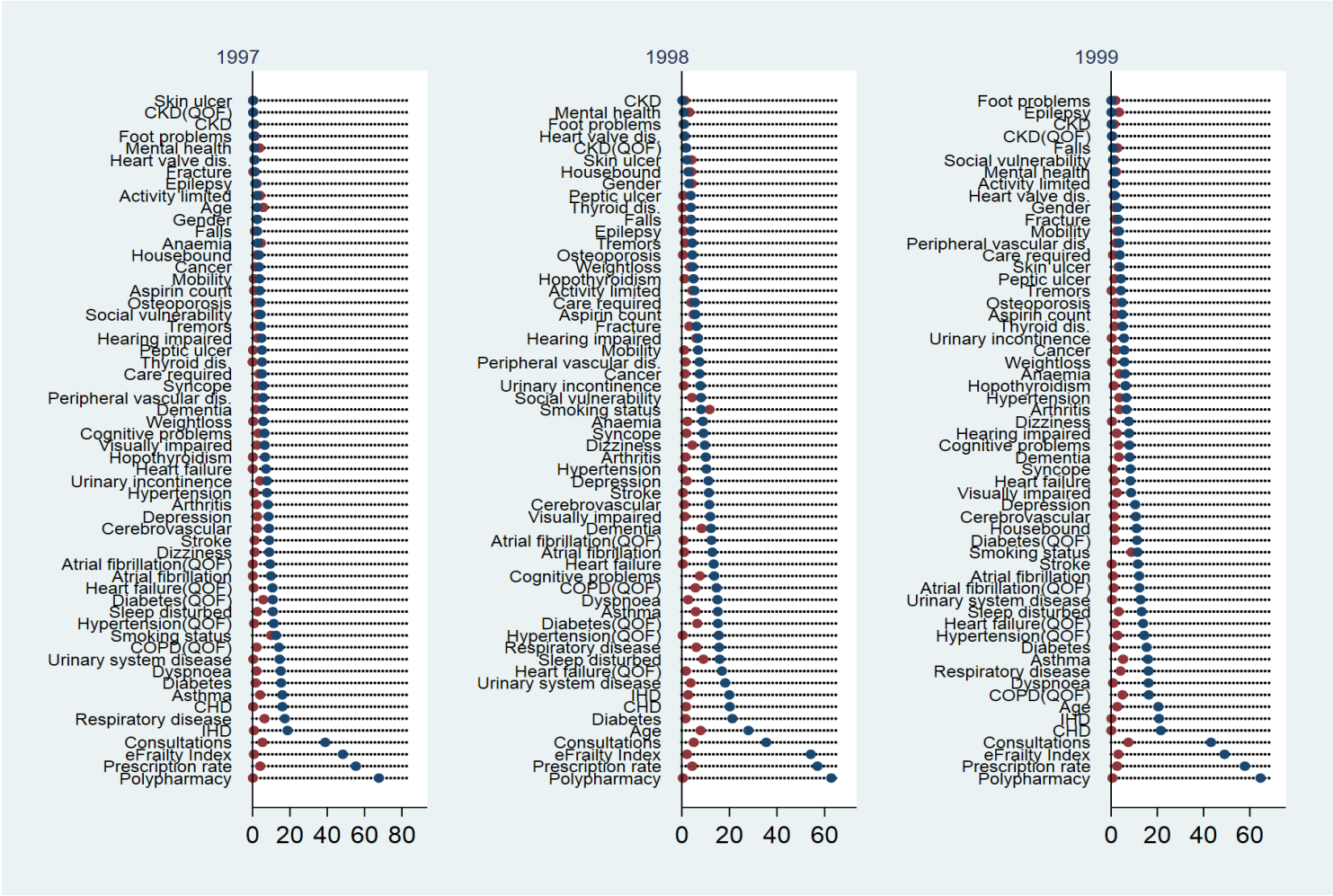


*Figure 34:* Plots of weighted (by propensity scores in red) and unweighted standardised (blue) mean differences for balancing variables – 1997 -1999 cohorts

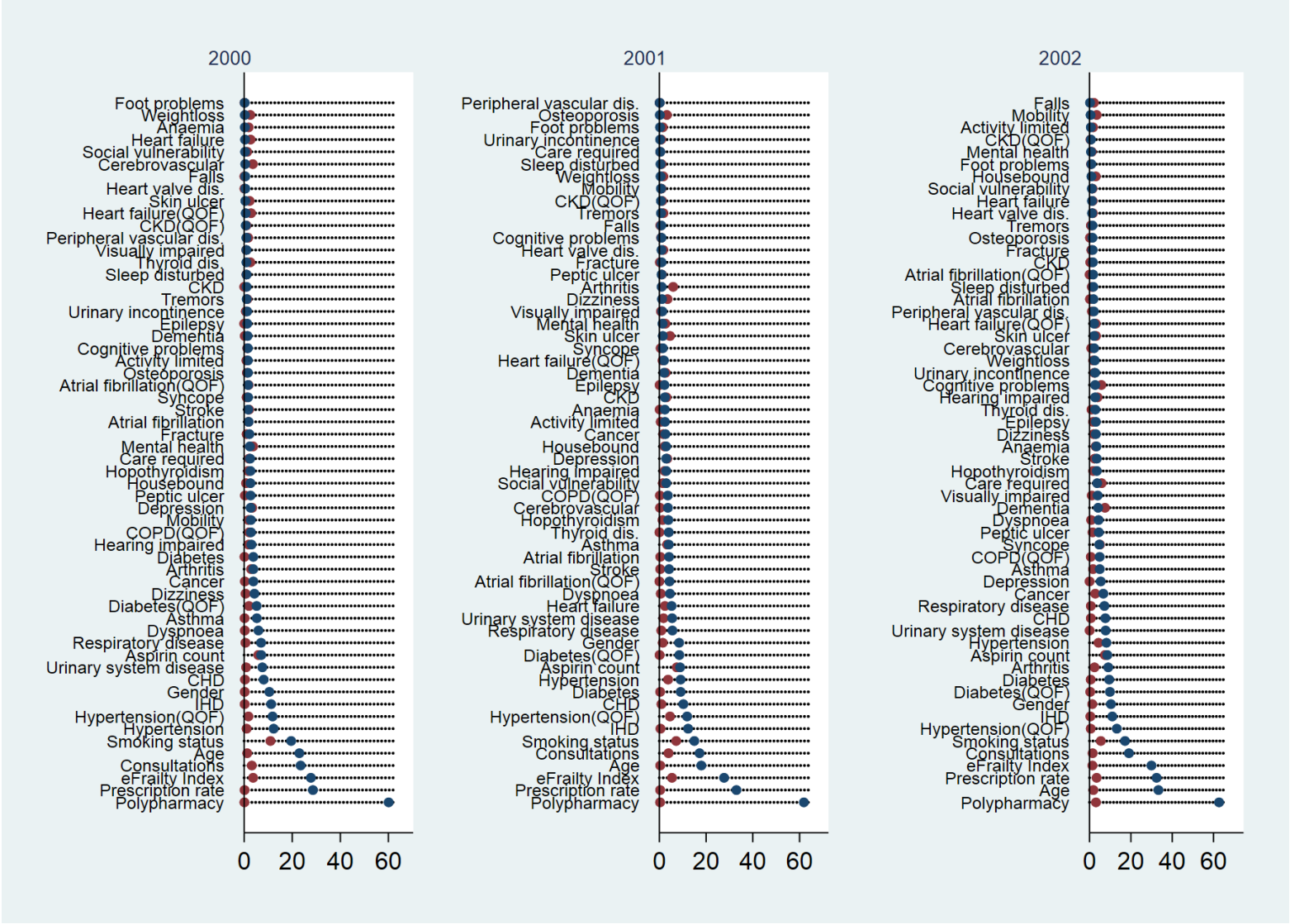


*Figure 35:* Plots of weighted (by propensity scores in red) and unweighted standardised (blue) mean differences for balancing variables – 2000 - 2002 cohorts

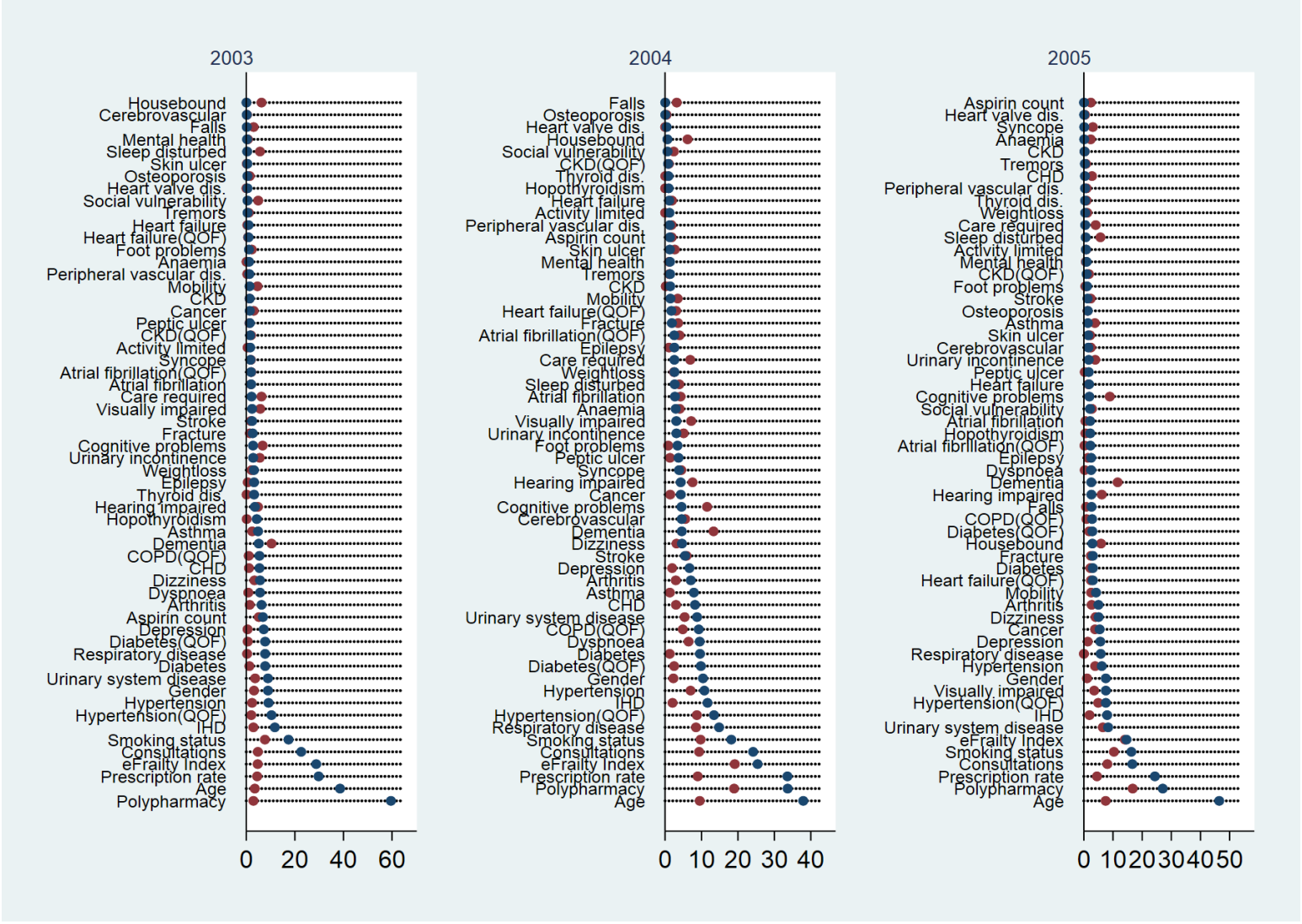


*Figure 36:* Plots of weighted (by propensity scores in red) and unweighted standardised (blue) mean differences for balancing variables – 2003 -2005 cohorts

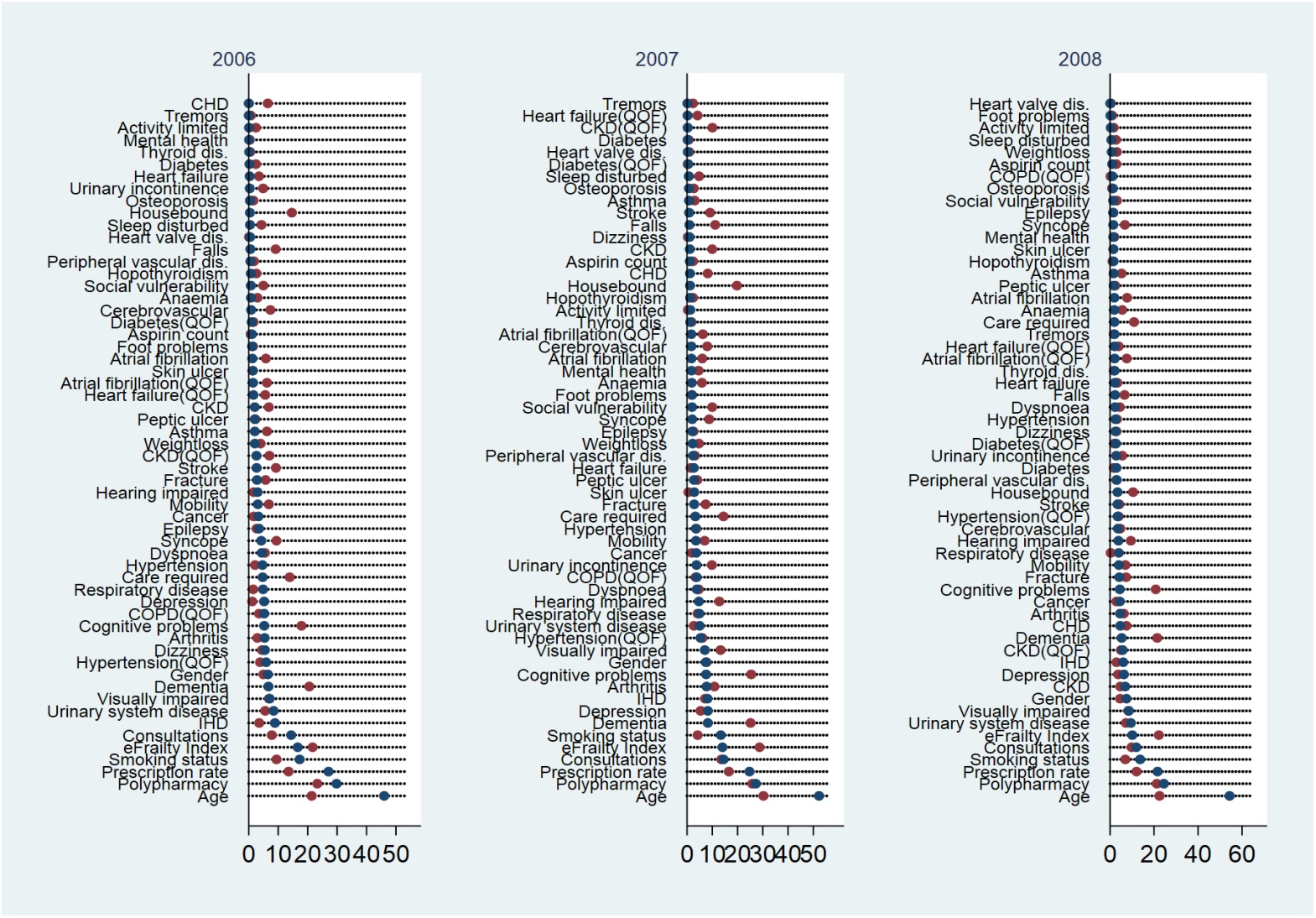


*Figure 37:* Plots of weighted (by propensity scores in red) and unweighted standardised (blue) mean differences for balancing variables – 2006 -2008 cohorts

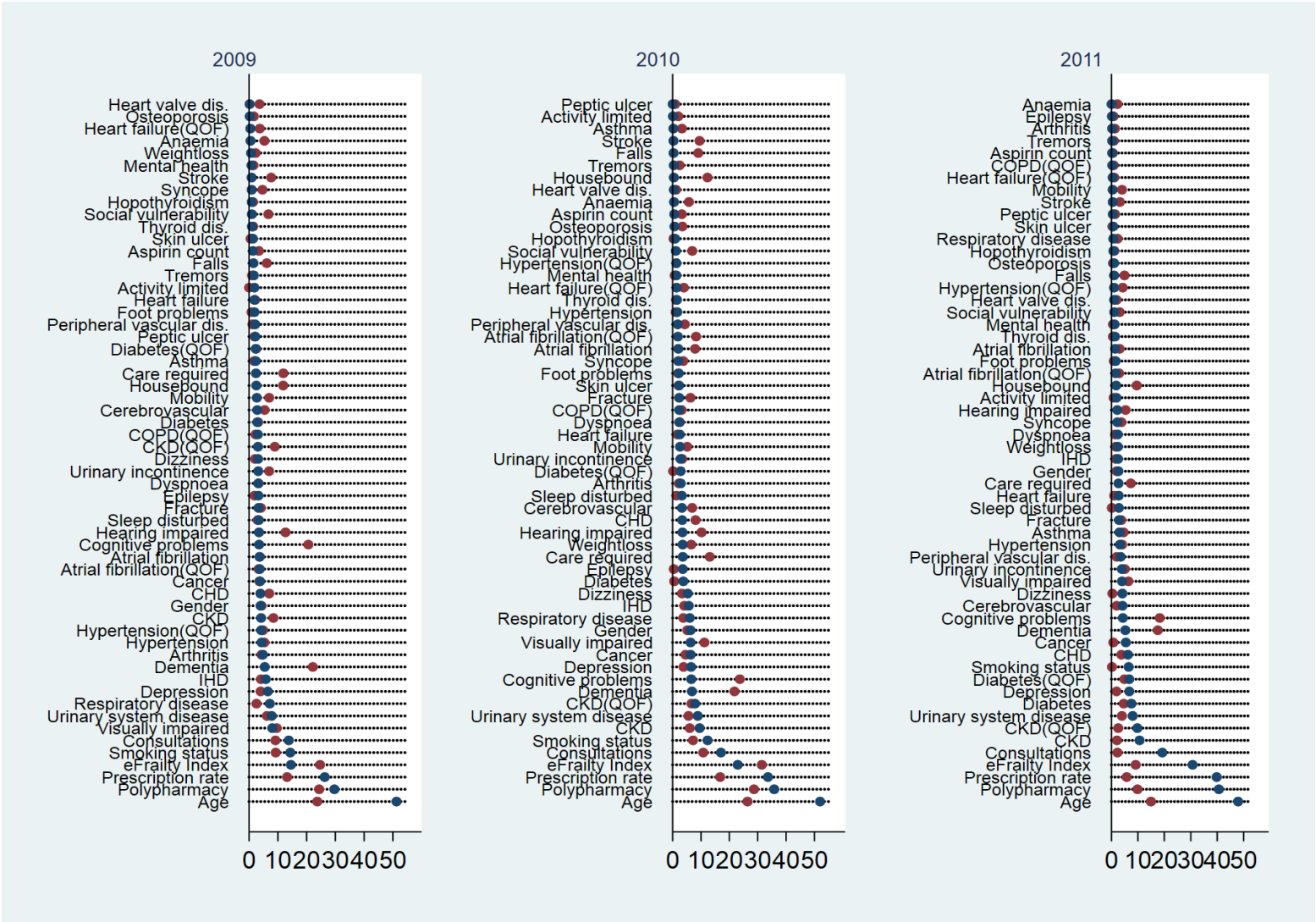


*Figure 38:* Plots of weighted (by propensity scores in red) and unweighted standardised (blue) mean differences for balancing variables – 2009 -2011 cohorts

# Chapter 7 - Discussion and conclusions

## 7.1 Summary of thesis

This project was partly motivated by the need for further evidence on vaccination in older adults from the routinely collected data stored in electronic health records. Vaccination of this age group is a key component of public health policy against infectious diseases, although studies have raised concerns that this age group is under-represented in clinical trials. Evidence from EHRs may be useful in supplementing existing evidence from RCTs, or even in informing the design of future trials. Therefore, in addtion to addressing the clinical question, this PhD necessarily focussed on the methods for deriving this evidence in the wider context of the growing interest in analysing routinely-collected data, such as electronic health records. However, in order to elicit inferential-quality evidence, such methods need to address the problem of confounding in observational data. Confounding is a problem for inference even from the “gold standard” data of clinical trials, where drop-out, non-compliance, poor randomisation, and even poor design, may lead to bias, if treatment groups become imbalanced in any prognostic factors. The bias may be corrected by modelling known confounders. However, this relies on having information about the relevant confounders, and in the absence of randomisation, it may not be reasonable to expect to identify and observe all potential confounders. This is none more so the case than with observational data, such as EHRs, where adjusted regression alone might not entirely account for unmeasured confounding. A greater focus is required on the use of quasi-experimental methods to deal with unmeasured confounding to better enable EHRs as a data resource in research. Acknowledging this issue was a major part of this project as I conducted a methodological review of QE methods in the context of EHRs and other similar data with longitudinal information on individual patients. This study, published in the Journal of Clinical Epidemiology, searched for published studies that applied QE methods to such data, and examined how each may have utilised the available, patient-level, longitudinal information to adjust for, or rather mitigate for, unmeasured confounding [322]. In this study, I aimed to draw attention to the different QE methods that leverage longitudinal information to adjust for confounding, some of which may

be underutilised due to the lag between development and widespread adoption in research.

To address the problem of confounding in vaccination studies, I focussed on the application of the prior event rate ratio method, and the framework of the Pairwise model. The PERR and Pairwise models have been an essential innovation in enabling the use of EHR data for replicating clinical trials studying survival and rate data. In this project, I applied these to address the clinical questions posed in the vaccination studies. I also reported on the relative performance of these methods to provide guidance for future work. The results from the PERR and Pairwise approaches may be compared to those from trials as a form of sensitivity analysis of the degree of unmeasured confounding. Therefore, as complementary and useful approach to the problem of confounding, the search strategy for the methodological review of chapter 2 was scoped to include sensitivity analysis too, the results for which were presented in the following chapter. As such, QE methods, like the PERR and Pairwise methods, may offer an alternative approach to sensitivity analysis and the issues over the transportability of external data on, and assumptions about, unmeasured confounders, by invoking a different set of assumptions to utilise longitudinal information within the same dataset.

Following the methodologial review of QE methods and sensitivity analyses as wider context for the PERR and Pairwise methods, I then conducted a review of these two methods, and presented a summary of their assumptions aggregated from different methodological studies, discussing their relative performance under different types of confounding bias. I also presented my literature review of studies that had utilised the PERR and Pairwise methods since publication of the seminal studies. As such, chapter four may offer further guidance to utilisation of the PERR and Pairwise methods, and will be submitted for publication.

Thereafter, the PhD focussed on applying the PERR and the recently developed Pairwise framework to the study of the real-world effectiveness of vaccination in older adults, an age group that may present many issues in the recruitment to trials . This was the first known example of the application of the Pairwise method outside the seminal work on this and the PERR-ALT method, and the first comparison of its performance relative to the PERR method in an applied setting. Moreover, the larger,

real-world data enabled analysis of interesting sub groups with surprising results. Both pneumococcal and influenza vaccines were studied, and the conditional estimates from the PERR and Pairwise methods were compared with the marginal estimates from a regression weighted by high-dimensional, propensity score based on observed confounders. This robust approach not only compared the success of adjustment, but served to gain further insight into the source and size of confounding bias, informing recommendations for good practice.

The investigation of pneumococcal vaccination presented somewhat controversial results in showing an increase in effectiveness with age. This runs counter to the hypothesis that immunosenescence may reduce effectiveness. However, while the degree of bias may be disputed, the direction of effect with age persisted to varying degrees in the results from all the methods used to address confounding. Whether or not immunosenescence leads to weaker immune response in older adults, or a shorter duration of immunity, the study found older adults to be at greater risk of infection. On that basis, vaccination still seems to confer immunity, with the real-world evidence suggests we should expect to see a net benefit from immunisation at the population level in older age groups in spite of immunosenescence. I presented these results at an oral session at the 33rd International Conference on Pharmacoepidemiology [323], and the report, which is included as chapter 5, is prioritised for submission to a suitable journal.

Complementing the work on pneumococcal vaccination, the project also investigated the effectiveness of the influenza vaccine as a secondary prevention against myocardial infarctions. The results for this were presented over two oral sessions at the 34th International Conference on Pharmacoepidemiology [324,325]. With influenza infection designated as a secondary outcome in this study, the study found consistent evidence for the effectiveness against MI with the influenza outcome as a possible mediator. However, gender was a much stronger prognostic variable for MI outcomes. Furthermore, while there was evidence that effectiveness against influenza waned with age, no moderating effect was found for MI. These results were confirmed by both the PERR and Pairwise methods, although the latter tended to be further from the null, estimating a greater protective effect from vaccination. In contrast to the QE methods, the adjusted survival models produced hazard ratios that were for most years greater than unity. This ran counter to expectation, and

would suggest that adjustment on observed confounders was incomplete. After application of the PERR method, the adjusted estimates closely agreed with those from the PERR-adjusted naïve Cox models.

## 7.2 Limitations

In the novel application of the PERR and Pairwise methods to vaccination studies, it was important to understand the limitations as well as the relative performance of each method, particularly in relation to the assumptions and conditions detailed in section 4.3 of chapter four. It was also important to acknowledge the limitations in the application to studying vaccine effectiveness to inform future work. Comparing the two methods together and against methods adjusting for measured confounding raised some wider methodological considerations and in some cases highlighted unresolved issues in the application of the PERR and Pairwise methods:

1. Limit to scope of method review: The scope was restricted to methods with a control arm as an analogue to RCTs. Furthermore, the review did not consider study designs, but rather it focussed on analytical methods for longitudinal, observational data, which in many cases, such as cohorts for before-and-after comparisons, may be a necessary pre-condtion for the application of the method, such as PERR or difference-in-differences. The focus on comparative studies meant excluding single-arm, self-controlled studies, although these were later considered as part of the specific review of the PERR and Pairwise methods in chapter four. The role of sensitivity analysis in dealing with unmeasured confounding was also reviewed, but in order to keep the report to a publishable size, this was abstracted into its own chapter.

2. Verification of stable confounding assumption: In the application of the PERR and Pairwise models, the stability between the prior and study periods of the net effect of confounding was emphasised as a necessary assumption in section 4.3. It is likely that if the influence of individual confounders changes between the periods then this would lead to a violation of the constant bias assumption. Although it is not possible to directly ascertain a violation of this assumption by hidden confounding, some clues may found by inspecting the stability of observed confounders under the assumption that changes in these may be

associated with changes in the unmeasured confounders. In the vaccine studies, the diagnostic tools used for the IPTW analysis were employed for this purpose, as well as for determining the success of weighting. Here, the weighted and unweighted mean differences in the measured confounders between exposure groups in the study period were compared to those in the prior period. As with vaccination studies, any change in the rank of the mean differences of the confounders over the annual cohorts and between the study and prior periods may be evidence of unstable confounding. In the pneumococcal vaccination study, the changing set of confounders between the prior and study periods instilled caution in the interpretation of the results. For the influenza-vaccine study, where there were 15 cohorts to consider, there was a notable shift in the ranked mean differences, with age being the largest following the introduction of the policy to vaccinate the elderly against influenza. Where anomalous changes in the results of a series of cohorts are considered over time, as with the influenza study, it may also be helpful to inspect the between-cohort variation in the confounders over a longer time period to make inferences about the stability of confounding, and potentially explain any incongruity.

3. Repeatability of outcomes: Following application of the PERR and Pairwise methods to investigating vaccine effectiveness, further clarification is needed about satisfying the basic condition for a repeatable outcome in order to apply these methods. Firstly, in the influenza vaccine study, there was a potential problem over the independence of multiple MI events in the same patient. If there was dependence between successive events then each MI may be determined by different confounding effects, or different confounders altogether. Stratifying survival analyses by different orders of the event was discussed as a way of mitigating against order dependence within each period, but this alone would not invoke repeatability between periods of the PERR and Pairwise methods as the strata in one period would necessarily be different from that in the other. Secondly, the independence, and therefore, repeatability of events may be threatened in the study of infectious diseases, where an outcome in the prior period of a patient provokes an immune response that carries over into the study period.

4. Prior period as estimate of confounding: Whilst the size of the group effect in the prior period of the pneumococcal and influenza studies was taken as an indication of the size of confounding bias, it was not possible to exactly estimate this in the presence of potential assumption violations such as missing, but otherwise balanced covariates in the Cox model [225]. Although conformance to the proportional hazards assumption was broadly satisfied through inspection of diagnostic plots, minor deviations would have also affect the precision of the estimates. While important prognostic variables are unlikely to be perfectly balanced in real-world data and the prior period will be a reasonable indicator of the pre-existing bias between levels of the exposure variable, some caution may be required in using this as an unbiased estimate of the confounding effect.

5. Determining successful adjustment against assumption violations: The vaccination studies offered an opportunity to re-examine the relative performance and limitations of the PERR & Pairwise methods using real data rather than simulated data. The models in both vaccination studies were adjusted for the common confounding variables of age and gender. However, simulation studies have demonstrated that bias can still affect the PERR method, if significant, yet balanced covariates are excluded from the model [225]. Therefore, the results from the PERR and Pairwise analyses were compared to IPTW regression. It was noted that the IPTW results were generally closer to the null. However, in the influenza vaccination study, many of these exceeded unity by a large degree, which prompted speculation about the correct specification of the model and terms used in the propensity score. Previous research has warned about model misspecification exacerbating the imbalance in other confounders and the use of propensity scores in practice [22]. Without full information regarding the unobserved confounders, it was not possible to distinguish between bias from model instability, model misspecification and unmeasured covariates, but by identifying as many measured confounders as possible, the degree of unmeasured confounding could be better ascertained. This exemplifies the challenge in using observational data. However, once the PERR method was applied to the IPTW results from the influenza study, these closely resembled the Pairwise and unweighted PERR results. From this it could be concluded that the PERR and Pairwise methods may have successfully corrected for residual

confounding error, and that any potential misspecification in the propensity score was either minor, or cancelled out by similarly misspecified prior periods. Best practice should therefore be to compare the QE results against the best available non-QE method that explicitly adjusts for measured confounders. The PERR method should then be applied to the adjusted results from the multivariable or weighted analysis. In this way, interpretation of the results needs to balance the potential problem of model misspecification against the assumptions of any applied QE methods.

6. Severe confounding between vaccines: Neither vaccination study was able to fully disentangle the effects from one vaccination in the presence of the other. This could certainly be explained in part by the size of the sub-groups. The pneumococcal study was divided into four sub-groups of influenza vaccination, classified by vaccination states in the prior and study periods. However, there were few patients in three of the four sub groups, with the attendant wider confidence intervals and uncertainty over the point estimates. In the influenza study, interest was in pneumococcal sub groups, dividing patients into never and ever having received the pneumococcal vaccine. However, there were far fewer patients, who had had a record of pneumococcal vaccination before 2002. This led not only to wider confidence intervals, but also spuriously high hazard ratios suggesting an unexpected harmful effect of the influenza vaccine in this sub group. It is possible that either by chance, confounding bias is less stable between the prior and study periods in a small sub group, or the sub group was an intrinsically less stable set of patients of rapidly increasing frailty. While EHR may offer the opportunity to study clinically interesting sub groups, no method can disentangle the effects of severely confounded exposures or treatments.

7. Season-specific outcomes: Specific to seasonal diseases, such as influenza, the approach in some studies has been to restrict analysis to outcomes that have occurred within a period defined as high pathogen circulation [270,326]. This would increase the specificity of the disease identification and reduce the inclusion of false positives in inferential studies. This may have been a reasonable strategy to follow, although in this project, the pneumococcal study was designed to span multiple years to maximise the collection of outcomes and so afford more

statistical power. The same approach was applied to the multiple cohorts of the influenza study, following up patients for a year rather than during individual flu seasons. Other before-and-after studies were noted in chapter six as having suggested the low-circulation periods outside of the influenza season as an adjustment period [318–320], and this was the strategy used in a recently published study [269] into influenza effectiveness using the PERR. However, choosing to adjust with an outcome from a low circulation period may be ill-conceived, as it could risk decreasing the specificity of the outcome since without laboratory confirmation, some cases may have been incorrectly identified. Also fewer outcomes would lead to less precision and wider confidence intervals. More importantly, the confounders affecting a low circulation study period are not necessarily the same as those affecting a high circulation prior period, and therefore may not offer a reliable basis for adjustment.

8. Pathogen evolution: While pneumococcus, like all bacterium, are known to evolve, particularly in response to treatment, vaccination and targeting by the immune system, constant surveillance by sentinel laboratories is required to respond to annual changes in the mix and type of influenza viruses, and update the vaccine accordingly. So, while the conclusions for the pneumococcal study remain reasonably robust to pathogen evolution over a short peroid of time, this was a potential weakness of the influenza study. Accounting for viral potency, antigenic drift and any subsequent mismatch between viruses and vaccine could improve the accuracy of the results, and potentially stabilise confounding between the prior and study periods. As noted in chapter six, operationalising an adjustment would be complex, and antigenic drift and viral activity would need to be quantified. In determining the effect on MI outcomes, this would also likely involve accounting for mediation by receptor proteins and potentially a differential cardioprotective effect of vaccine strains.

9. Viable alternatives: The scope of this PhD project was limited to a review and application of the PERR and Pairwise methods in the context of other QE methods. While this precluded focussing on other possible methods, such as those reviewed in chapter two, these should not be entirely overlooked as viable, alternative analytic methods for addressing the clinical questions of chapters five

and six. The test negative case control design is popular in the monitoring of influenza vaccine effectiveness [251,252,254,327]. This seeks to minimise confounding from health-seeking behaviour through restriction to those patients, who have been diagnosed as having an influenza-like illness (ILI) [255]. Cases of influenza are then confirmed among the ILI cohort by laboratory analysis, which facilitates the case-control analysis. However, generalisability of the results from such an analysis is arguably diminished by the restriction limiting the study cohort to ILI cases and the availability of laboratory testing to confirm cases. Because of the reliance on restriction, this method did not qualify as a QE method in the method review, but a review of its relative performance would be informative.

10. Survival data vs. rates: Although the focus in the project was on specific outcomes expressed as survival data, it is worth mentioning that the PERR could have been applied to events as rates as a Poisson-like process. This, however, would have required parametric modelling and would have lacked the flexibility of the Cox models.

11. Estimation of time-varying covariates: The expression of vaccination in the Cox model as time-varying covariate was found to greatly facilitate the derivation of survival times for the cohort. Without this tool, the survival start times would start relative to the date of vaccination for the vaccine-recipients. However, to avoid disparity in the distribution of start times with the controls, who would otherwise commence from the same index date, it was necessary to map the vaccination dates onto the controls as their own start dates. To avoid unintentionally introducing further bias, this entailed writing a computationally intensive matching algorithm in the statistical software. Once the start times had been designated for the controls, the matching was no longer needed. Expression of the vaccination variable as a time-varying covariate obviated the need for matching, since both vaccine-recipients and controls started their follow-up from a common index date. With the extra pre-vaccination period for vaccine recipients in the study period, the point estimates seemed to be less stable over time and the confidence intervals were slightly inflated above those for the PERR and Pairwise estimates.

12. External validity/generalisability vs. selection criteria: Although one of the much cited advantages of using RCD over RCTs is the generalisability of the results to a wider population, some of the selection criteria used in applying the PERR method may have placed restrictions on the external validity of the findings. Some of the settings were the exclusion of patients based on previous vaccinations, and the exclusion of controls, whose follow-up may have included vaccination during the next vaccination season.

## 7.3 Discussion of thesis

At the beginning of this thesis, some time was spent setting the scene and describing the motivation for this project that had arisen from the growing interest in using routinely-collected data for clinical evidence. In spite of the challenges of deriving plausible evidence of causal relationships using such data, this interest has not diminished. Once the preserve of commercial interests and data mining in market research, the presence of big data in medicine is only likely to grow. A systematic review of methods for dealing with the problem of confounding was not only an obvious means of providing the necessary background required for this PhD project, but also served as a timely reference for other studies eliciting causal inference. For this purpose, the review was directed towards the application of EHRs for causal inference, with a specific interest in utilising the longitudinal information available for each patient. In this way, the PhD project would contribute to the growing interest in using quasi-experimental methods [322]

The methodological review of chapter two provided an essential broader context for the methods among which the PERR and Pairwise sit. It was also noted that many of the QE methods have their roots in econometrics, but that uptake of these methods in medical statistics might not entirely match the growing interest in EHRs. It would be interesting to ascertain the relative frequency of QE methods used to analyse EHRs against explicit adjustment methods that assume no unmeasured confounders, although this may involve undertaking a much larger review. However, it should not be assumed that conventional multivariable regression, matching or weighting will fully adjust for unmeasured confounding bias that has previously been described as “stubborn” and resistant to adjustment in EHR data [22,213], including

vaccination studies [318]. This justified the recommendation in the conclusions of the method review to utilise the longitudinal information available in EHRs, where possible, as good practice when seeking causal inference from this data. Since causation exists as a temporal relationship, a longitudinal dimension is an often undeclared condition in the theory of causal inference that was discussed in the introductory chapter. Therefore, researchers using EHRs need to be thinking longitudinally about the threats to unbiased causal inference from unmeasured predictors of an outcome that may exist as part of a wider causal network, potentially confounded by past exposures and outcomes. For that reason, it makes sense to think longitudinally about adjusting for unmeasured confounders too. The method review of chapter two, the sensitivity analyses covered in chapter three and the discussion around the PERR and Pairwise methods in chapter four offer many strategies for dealing with unmeasured confounding and are presented as a collection of resources for improving causal inference from longitudinal, observational data.

Faced with untestable assumptions about unmeasured confounders, it should be considered good practice to supplement adjustments for measured confounders with a sensitivity analysis, QE analysis, or QE adjustment based around a particular study design. A QE method may also be deployed as sensitivity analysis to test for the presence of no unmeasured confounders. In this respect, although each QE method invokes certain assumptions, QE methods generally may offer an advantage over traditional SA methods, which often rely on the transportability of external information about confounders and their applicability to the dataset under analysis. In contrast, the QE before-and-after methods, like the PERR and Pairwise, utilise within-dataset, longitudinal information.

Where unmeasured confounding is readily acknowledged, and the primary analysis relies on a QE adjustment, then an analysis based on an explicit adjustment for observed confounders may be considered a sensitivity analysis and a check on the assumptions of the QE method. This robust approach was followed for vaccination studies with the PERR and Pairwise results firstly compared with each other. The underlying study and prior period  estimates of the PERR estimate may also diagnose the existance of bias. Covariates in the basic adjustment comprised the main confounders of gender and age, both of which are known to be main

determinants of frailty in the older population, and as such may partly account for other associated sources of bias. Contingent on the assumptions for the Cox model being met, the vaccine was expected to show no effect in the period prior to vaccination, so the size of deviation from the null in the hazard ratios for vaccination was interpreted as a measure of bias in the data due to unmeasured confounders. In Lin and Henley [192], a procedure was presented for testing for the presence of unmeasured confounding in the prior period and to differentiate this from imbalanced censoring. However, in the prior periods of the vaccination studies, survival times were administratively censored at the end of follow-up in the absence of an event. HRs for vaccination significantly greater than unity in the study period were considered inconsistent with existing evidence, and therefore implausible. Since selection bias in the vaccine studies was directed towards the frail, rather than exhibiting a healthy user bias, the degree to which the study period exceeded the null was also interpreted as an indicator for the presence of unmeasured confounding bias and its potential size.

Finally, the results from the PERR and Pairwise methods were compared with those from an IPTW analysis, with weights based on propensity scores predicted from all observed variables determined to be confounders. The IPTW estimates offered a more easily interpretable marginal estimate of effectiveness unconditioned on any other variable, although this precluded estimation of interaction terms, instead relying on sub-group analysis. The direction and size of deviation of the IPTW HR from unity for the exposed group in the prior period could be informative about the size of residual confounding or at least indicate its potential presence in the study period. This is of course contingent on the propensity score model being correctly specified and on compliance with the proportional hazards assumption of the Cox model. The exposure HR for the prior period should then be interpreted as the bias having averaged over the effect of the observed confounders.

As a further verification of the PERR and Pairwise results applied to an unadjusted or partially adjusted multivariable model, the PERR method could also be applied to the weighted estimates of the study period using weighted estimates from the prior period. In the pneumococcal study, some of the weighted prior-period estimates were greater than unity indicating the possible presence of residual confounding, although instability in the confounders between the periods could not be ruled out.

However, in the study on influenza vaccine effectiveness on MI outcomes, many of the weighted study period HRs exceeded unity. This suggested the potential presence of residual confounding due to unobserved and unadjusted confounders, although some proportion of the bias could still have been attributable to model misspecification. After applying the PERR adjustment to the weighted Cox estimates, the resulting point estimates were remarkably close to the PERR-adjusted estimates from the age and gender adjusted Cox model. This was a reassuring check on the performance of PERR demonstrating its ability to adjust to results that are consistent with each other, notwithstanding potentially misspecified, marginal and basic-adjustment models.

One of the challenges in applying the PERR and Pairwise to vaccination studies was the derivation of the survival start dates for the controls from the vaccine recipients through matching. Expressing vaccination as a time-varying covariate obviated the need for this, and the PERR was successfully applied to the influenza study with the vaccine effect as a time-varying covariate in the study period, although the confidence intervals were wider than the standard approach. Nonetheless modelling time-varying covariates is a highly adaptable approach to investigating comparative effectiveness studies, not just for accounting for time-varying effects, but also where, as in the vaccination studies, determining the start times for follow-up may involve complicated matching algorithms.

Throughout the vaccination studies, the Pairwise estimates were consistently further from the null than the PERR estimates. This may have been due to missing, but otherwise balanced covariates, which can bias the PERR method, but not the Pairwise. However, other differences exist between the performance of each method under the same assumption, so it was difficult to determine the provenance of the differences. For instance, where there is an interaction between exposure and a hidden confounder, then the resulting bias has been shown to vary according to size of the confounding effect, although for most of the range in one scenario presented in Lin and Henley [192], this was shown to favour the PERR method. Further bias due to assumption violation may arise when the exposure or treatment is indicated by the event in the prior period. Here, the relationship between the bias and the strength of indication for each method differentially varies over the range or the effect of confounding. As discussed in paragraph 4 of 4.3 of chapter four, this is further

complicated by indication existing on a continuum rather than as a binary state. Furthermore, trying to diagnose confounding by indication may be difficult, given there would be nothing to differentiate between any treatment that may have been indicated by a prior event and those treatments that might have been administered coincidentally following an event (without actually being indicated by it). For instance, a patient could easily have had a chance pneumococcal infection in the period leading up to a routine appointment for vaccination against the disease, which may have appeared to have been indicated, but otherwise lacking any direct causal connection.

This project has demonstrated the PERR and Pairwise framework to be more than just a QE method applied to unadjusted models to mitigate against unmeasured confounding. Rather, such methods could be an integral part of a best-practice approach to EHRs, especially when combined with the best available adjustment for observed confounders. Firstly, the direction and degree of unmeasured confounding bias may be inferred from the prior and study periods contributing to the PERR analysis, contingent on a correctly specified model for those periods. Where multiple, successive cohorts are analysed, as was done for the study of influenza effectiveness on MI outcomes, then inference about the stability of the confounders may be inferred from the year-to-year changes in the prior periods, since only confounders and unmeasured covariates will be solely affecting the estimates. For instance, relative to the exposure effect, the effects of confounding may be deemed unstable, if the year-to-year variation in the prior period seems excessive over time. Comparisons of the unweighted mean differences in confounders between exposure groups from each cohort, as found in the diagnostic plots for the IPTW method, may further shed light on the stability of the unmeasured component of confounding bias through potential associations with the measured confounders (the principle upon which perturbation analysis sits [80]). Further comparison against methods attempting a full, explicit adjustment may help identify the proportion of bias that may be due to unmeasured confounding. However, caution is required as the accuracy of the results from the adjusted model will depend on the functional form of the adjusted variables and any potential interactions being correctly specified in the model.

As was demonstrated in the study of PPV effectiveness, the PERR and Pairwise methods could potentially provide unbiased estimates of interactions and exposure

modifiers. The investigation of effect modification is often proposed as a secondary analysis, requiring division of the data into sub-groups of the effect modifier. However, this may lead to challenges of statistical multiplicity, and many statistical corrections are either overly conservative, producing type 2, false-negative conclusions, or too liberal leading to type 1, false-positive errors[328]. The procedure followed in this project was to model the interaction between exposure and potential modifier, as a global test of the sub-groups. The interaction, when found to be significant in the study period, was then adjusted using the PERR method, which as far as I am aware from the published literature, is the first example of deploying the PERR adjustment in this way. The need for further research into this particular application is acknowledged in section 7.4.3.1 below. The interaction could be interpreted directly through its PERR-adjusted effect, although interpretation for effectiveness at specific ages is complicated by also having to consider the PERR-adjusted main effects and interaction simultaneously. Therefore, when the global test of the age and vaccination interaction was found to be significant in the 2005 cohort of the PPV study, the effect of age was interpreted through modelling three main effects for in the data, having divided it into age sub-groups. An advantage of this approach is that the marginal effect of vaccination could be estimated using IPTW for each age group for ease of interpretation. Furthermore, the PERR method was also applied to adjust for any residual confounding not controlled for by the propensity score used in the IPTW method.

The authors of the seminal work on PERR, and subsequently PERR-ALT, applied the methods to replicated trials using EHRs as a means of validation, and comparison to the trial results. However, as reported in chapter four, the application of PERR varied in the subsequent studies. While few applied PERR to complete the adjustment for confounding, most applied the method to unadjusted models as a means of presumably adjusting for all confounding. Here, this particular chapter will be published as a review with a view to offering guidance and best practice in the application of the methods. A cautious approach was taken in this project, applying the PERR and Pairwise methods to models conditioned on the basic adjustment for age and gender. These were compared to models adjusted for measured confounders through weighting, although the two approaches offer different estimates of effect, one conditional on age and gender, and the other, marginal,

averaged across all the other effect effects and characteristics of the sample. Finally, the marginal models were adjusted with PERR to gauge the degree of bias due to unmeasured confounders. This robust approach was in contrast to many other studies applying the PERR method. The novelty of this work was also in the first ever application of the Pairwise, and PERR-ALT, model outside of the seminal studies.

A notable difference in the approach to the application of the PERR method between the seminal studies of Tannen, Weiner et al and this project, was that here the methods were applied to the vaccine data without attempting to exactly replicate RCTs, and did not specifically consider an as-treated analysis. Without the tool of randomisation, one would reasonably expect the replication of a trial in RCD would produced biased results, but replication has been used primarily to help validate the PERR method in the seminal work. Hence, any differences between the replicated and original trials results could be attributed to bias rather than trial conditions with all other things, apart from randomisation, being equal. Trial replication was not repeated for the vaccine studies as the purpose was not to limit the generalisability and sample size by imposition of too many selection criteria. While some exclusion criteria may serve to hone a clinical question to a relevant population, the applicability of results from real-life practice relies in part on studying exposures away from carefully controlled trial conditions. With regards to the sample set, an as-treated approach was unnecessary given that exposed individuals were necessarily compliant for the duration of the vaccination effect, unless they were censored on death or upon deregistering with their practice. In the case of influenza, the duration of vaccine effectiveness was for the year from the point of vaccination, while for PPV, which is not administered annually, the duration of effect was considered to be long enough to accommodate a follow-up of at least two years.

Evidence for moderating effect of age on the effectiveness of the pneumococcal vaccination would appear to be reasonably robust given that this was observed in the naïve Cox model adjusting for age (a main effect of the interaction term) and gender; the marginal effects of the high-dimensional IPTW model by age group; and the PERR and Pairwise models applied to the former. The difference between the age groups of the IPTW model was less than that suggested by the interaction terms of the conditional models, but this may have been attenuation of an otherwise large interaction effect by residual confounding. The accuracy of the PERR and Pairwise

adjusted estimates still rested on the assumption of stable confounding, but these were arguably less exposed to changes in pathogen virulence than in the study of the influenza vaccine. Further simulation work may be needed to test the performance of the PERR and Pairwise methods in estimating interactions (see future work below). However the findings of the age-vaccine interaction could impact on the direction of future research [70], with the real-world evidence suggesting that the benefits of pneumococcal vaccination are not noticeably undermined by a poorer immune response in older ages.

An important implication from applying the PERR and Pairwise methods to vaccine studies is that these methods may be more suited to studying non-infectious diseases, or adverse events where the outcomes are more likely to be independent and repeatable. However, the challenges of applying the methods to these studies allowed a thorough comparison of their performance using real-world data. In spite of these challenges, the evidence from the vaccination studies were supportive of their protective effect. For PPV, effectiveness increased with age, and for influenza, the protective effect against MI was mediated through its effectiveness against influenza. Furthermore, the results for the influenza vaccine aggregated over many cohorts suggest an effectiveness of 39% against MI, which was consistent with the evidence from one of the few trials conducted on this outcome.

Lastly, while the results from this analysis of EHRs are compared to trial results, the evidence from real-world data should not necessarily be viewed as a challenge to RCT-generated evidence. Instead, real-world evidence tries to acknowledge the complexity of the greater causal network, in which the evidence from RCTs might sit. The findings from this project have not only contributed useful clinical evidence, but have also demonstrated the utility of routinely-collected data, as well as contributing to the research into methods that can analyse such data by mitigating for, or describing the sensitivity to unmeasured confouding.

## 7.4 Future research

Future research that may progress the work undertaken in this PhD project can be described as two-pronged. While interest may lie in further validating and extending PERR and Pairwise as viable methods in research using routinely collected data,

there is also a need to address some pressing clinical questions around drug-vaccine interactions, which may be resolved by the advantages afforded by the real-world data from EHRs. One of these questions follows on from discussions at the 33rd International Conference on Pharmacoepidemiology prompted by my work on pneumococcal vaccination in this PhD project. Some pilot work has been undertaken to explore the viability of such a project, and more details are given in the protocol for this, which is presented below in section 7.4.1. In addition to the application of the methods in further clinical research, I identified during the course of this project, further ideas for developing and extending the methods.

## 7.4.1 Future research: A protocol for the investigation of the influence of statins on influenza vaccine effectiveness

### *7.4.1.1 Background*

An elevated risk of acute myocardial infarction has frequently been observed to coincide with high levels of influenza infection [302,304,305,329,330]. It is believed this is a likely inflammatory response to infection and to the release of proinflammatory cytokines. Some studies have also found evidence of a reduction in the rates of myocardial infarction and heart failure following influenza vaccination [311,314,331–333]. This, in turn, this has led to a proposed additional role for influenza vaccination in preventing acute cardiovascular diseases [306,316,334–336], although some of the observational evidence may be biased without adjustment for unmeasured confounding [311,314]. Furthermore, estimation of the marginal effect may conceal differential effects between key sub-groups of patients, with some patients deriving great benefit from vaccination [332,337]. Because statins may have anti-inflammatory as well as lipid-lowering effects, statins have been proposed as a possible prophylaxis against influenza as well as cardiovascular diseases on the basis of the proinflammatory stimuli, common to both diseases [338]. However, recently published evidence has suggested that statins may impair immunogenicity, and therefore effectiveness, of the influenza vaccine [339–342].

Building on previous work investigating pneumococcal vaccine effectiveness and the role of the influenza vaccine in protecting against myocardial infarctions in the elderly, this study proposes to investigate whether statins moderate vaccine effectiveness against influenza in the population of adults aged at least 65y.

Although the US data are subject to different biases and different prescribing patterns, to facilitate comparability, the settings for this project will largely be informed by the study of Izuerieta et al.[342]. By replicating those studies conducted on predominantly US populations, this project aims to uniquely contribute to the growing body of evidence on the immunomodulatory effect of statins using UK population data. Furthermore, while previous studies have adjusted for observed confounders, we will also investigate what role unmeasured confounding may have by using quasi-experimental techniques[322].

#### *7.4.1.2 Cohort selection*

This project will use data collected on UK adults aged at least 65y available from the Clinical Practice Research Datalink. The study will start from 2002, avoiding the instability in treatment and potential confounders during the change from the risk-based practice to an age-based vaccination policy, which completed in 2001. Annual cohorts will then be recruited to the latest annual cohort that can be guaranteed complete and up-to-standard by the CPRD. Recruitment will begin on the index date for each year, arbitrarily set to the 1st September as the date likely to precede seasonal vaccination (typically October) and the maximum number of patients recruited in the same period. Patients will be included if they have a record of vaccination against influenza between the index date and the 31st January of the following year. Follow-up will begin on the date of influenza vaccination, plus 14 days to allow for full immunogenicity, and end 30th April.

For the purpose of applying a quasi-experimental adjustment for unmeasured confounding, a sub-cohort of patients will also be selected according to their vaccination status in the prior period, defined as the 1st September of the previous year to the 31st of the following January, during which follow-up will last until 30th April. In this way, the sub-cohort will be defined as patients, who have been vaccinated against influenza for two consecutive years. For example: the index date for the 2010 cohort would be defined as the 1st September 2010, and patients required to have been vaccinated between the index date and 31st January 2011. The study period would therefore be defined as the follow-up from the date of vaccination to 30th April 2011. Follow-up in the prior period would have ended on 30th April 2010 and begun with an influenza vaccination in the prior period between 1st September 2009 and 31st January 2010.

Patients should be HES-linked, alive and at least 65y of age on the index date, and excluded if not registered at their practice at least two years before. The General Practices will also have to be up-to-standard at least two years before the index date. An absence of any clinical consultation in the five years before the index date will be regarded as unlikely for patients in this age group and so such cases will be excluded. Instances of influenza vaccination will be determined by their relevant medical codes in the immunisation file and product codes in the therapy file (Appendix D – CPRD and HES codes)

#### *7.4.1.3 Intervention*

Statins listed under the relevant British National Formulary chapter code of 0212 can be identified in the CPRD therapy file by their product codes (all with CPRD BNF code = 02120400). Synthetic statins are Atorvastatin, Rosuvastatin and Fluvastatin. Non-synthetics are Simvastatin, both with and without Ezetimibe, and Pravastatin (see Appendix F – codes for statins in CPRD data, for full list of statins and product codes). For any given cohort, records of statin prescribing will be retrieved for the period between 30th April of the year before the index date and 16th January of the year after the index date (e.g: for the 2010 cohort, this would be from 30th April 2009 to 16th January 2011). The end date for a prescribed course of statins can be calculated from the quantity supplied and the date of prescription. Courses will be counted as one, where prescriptions were renewed within two days of the previous one finishing. Patients will be excluded from the cohort if prescribed a spurious quantity of statins considered to be less than 7 or greater than 84. To ensure a statin-free prior period for the quasi-experimental adjustment of unmeasured confounders, patients will then be excluded if there is evidence of a statin prescription in prior period up to 15 days after vaccination in the prior period, and since the 30th April of the year preceding the index date. For example, given an index date of 1st September 2010, this would define a statin-free period from 30th April 2009 to 15 days after an influenza vaccination in the period between 1st September 2009 and 15th January 2010.

Statin users shall be defined as patients with a sufficient supply of statins to cover the period 15 days before vaccination through to 15 days after vaccination for each study period, with a maximum allowable gap of 2 days. Adherent statin users will be those who are prescribed statins in the period from 30th April preceding vaccination

in the study period to 15 days after vaccination with a medication possession ratio (MPR) of ≥ 0.8. Non-users will be defined as beneficiaries with no evidence of receiving statins during the period extending from 6 months before to 15 days after vaccination. Patients who do not meet the criteria for either group will be excluded

#### *7.4.1.4 Outcomes*

Given the rarity of instances of antiviral drug prescribing in the dataset, the study will examine the effect on two outcomes: hospitalisation for influenza and a composite of a prescription for an antiviral drug or hospitalisation for influenza. As a sensitivity analysis, eligibility of the outcomes will also be restricted to those occurring during periods of high level influenza circulation periods as confirmed by PHE laboratory surveillance. Survival times will be calculated as the time from vaccination until date of one of the events of the composite outcome, less the 14 day period for developing immunity after vaccination. Influenza diagnoses in the HES data can be identified by their ICD-10 codes (Appendix D – CPRD and HES codes), and the date of influenza given by the date of admission to hospital. Antiviral drugs are identified by their product codes (Appendix D – CPRD and HES codes).

#### *7.4.1.5 Statistical analysis*

The effect of statins on survival times until the composite outcome will be analysed using Cox's regression adjusting for age and gender, censoring on death. Any patients found to have vaccination dates occurring after their date of death will be dropped from the cohort. Any negative survival times resulting from the addition of 14 days to the vaccination date (reflecting the period for immunogenicity) will be assigned zero times, with any corresponding events being coded as right-censored events.

#### *7.4.1.6 Adjustment of cohorts for unmeasured confounding*

To adjust for unmeasured confounding, the resulting hazard ratio from the study period for the sub-cohort of patients with two years of influenza vaccination will be adjusted with the hazard ratio for statin-free prior period, using the prior event rate ratio (PERR) method. Confidence intervals for the PERR-adjusted hazard ratio can be obtained through bootstrapped resampling.

#### *7.4.1.7 Weighted analysis of 2010 cohort*

An inverse probability treatment weighted (IPTW) survival analysis will be performed to estimate the marginal effect of the statins, adjusting for measured confounders for the study period only. Stabilised IPTWs will be based on the propensity scores derived from a predictive model for statin treatment. Prognostic variables considered for the propensity score model will be those found to be significant at the 5% level in a Cox regression of the defined outcomes. Candidate variables will include age, gender, diseases recorded for the Quality Outcomes Framework [295] and symptoms and diseases derived for the electronic frailty index [292], as well as the index itself. To gauge the extent of unmeasured confounding, the IPTW estimates from the study period will be obtained for the sub-cohort of patients with two years of influenza vaccination, and additionally the PERR adjustment shall be applied.

### 7.4.2 Future research: Development of the post event rate ratio method

Investigation of vaccination rates during the recruitment to the pneumococcal vaccine study prompted an exploration of an alternative approach to the PERR method. The pneumococcal vaccine had been introduced into adults aged at least 65y by age group over three successive years between 2003 and 2005. Vaccination rates before 2003 according to the data were below 30%. This was consistent with the data from the HPA, who reported a vaccination rate of 29% in 2003 based on monitoring over the previous 10y. According to the CPRD data, this had increased to 72% in 2008, compared to 69% reported by the HPA for that year.

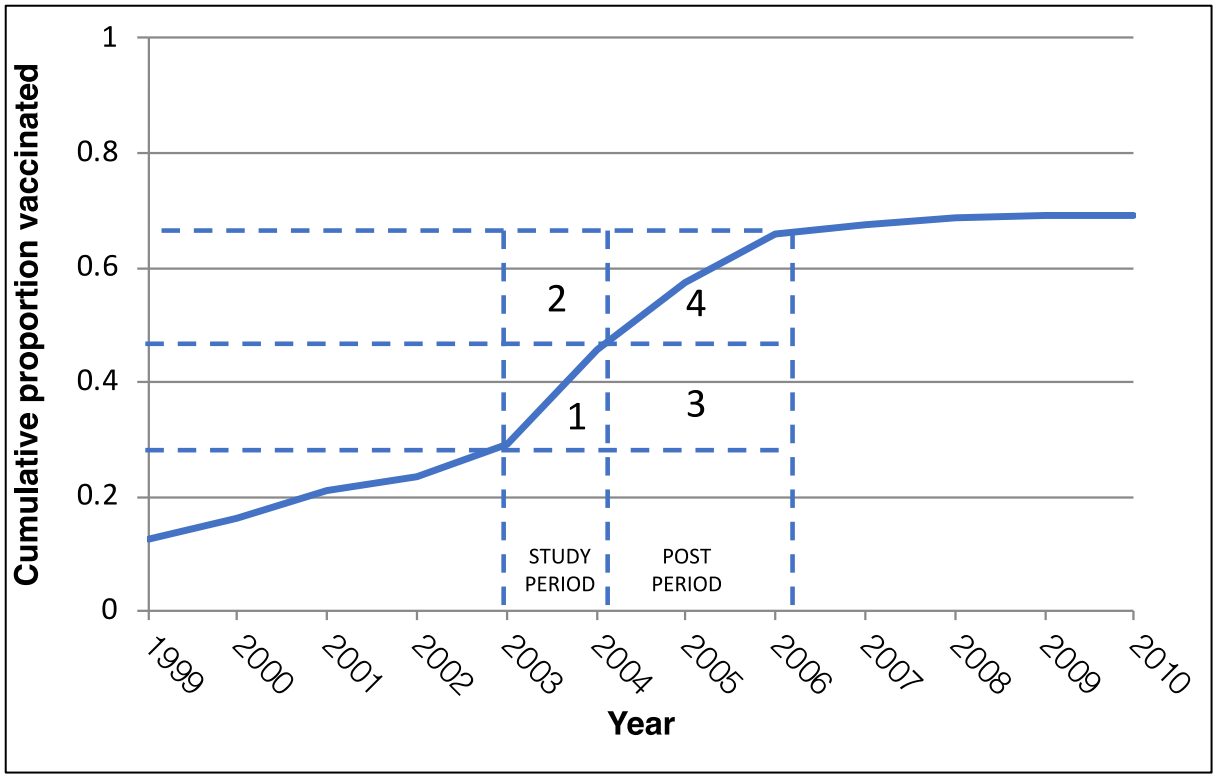


*Figure 39:* Plot of proportion of patients ever receiving a pneumococcal vaccination for each year in the CPRD data from 1999 to 2010, to illustrate potential for an adjustment of the study period using the all-vaccinated patients in the post period. Vaccinated patients up to the study start are excluded. The hazard ratio of treatment in the study period is estimated from the vaccinated group's survival times (found in area 1 below the curve in study period) relative to those of the controls (dotted quadrant marked 2). The hazard ratio of treatment in the post period is estimated from the vaccinated group's survival times (dotted quadrant marked 3) relative to those of the controls receiving the vaccine (found in area 4 below the curve in the post period).

With the PPV vaccination policy achieving its aim of increasing vaccination coverage, a large part of the UK population of older adults had rapidly transitioned from an unvaccinated to a vaccinated state (Figure 39). This prompted the idea to use the event rate from a "post" period after the study period, in which all patients in the study period were eventually vaccinated. The confounding adjustment would come from the post period, during which all patients would have been vaccinated, and the only observed effect should have been from confounding bias. The post period could then replace the prior period as an adjustment ratio for the study period HR (or relative risks if using rate data instead of survival). The proposed idea is as simple and intuitive as the original PERR method. It is proposed that this may be tested on the vaccine effectiveness data and compared with the results from the PERR and Pairwise methods (Figure 40).

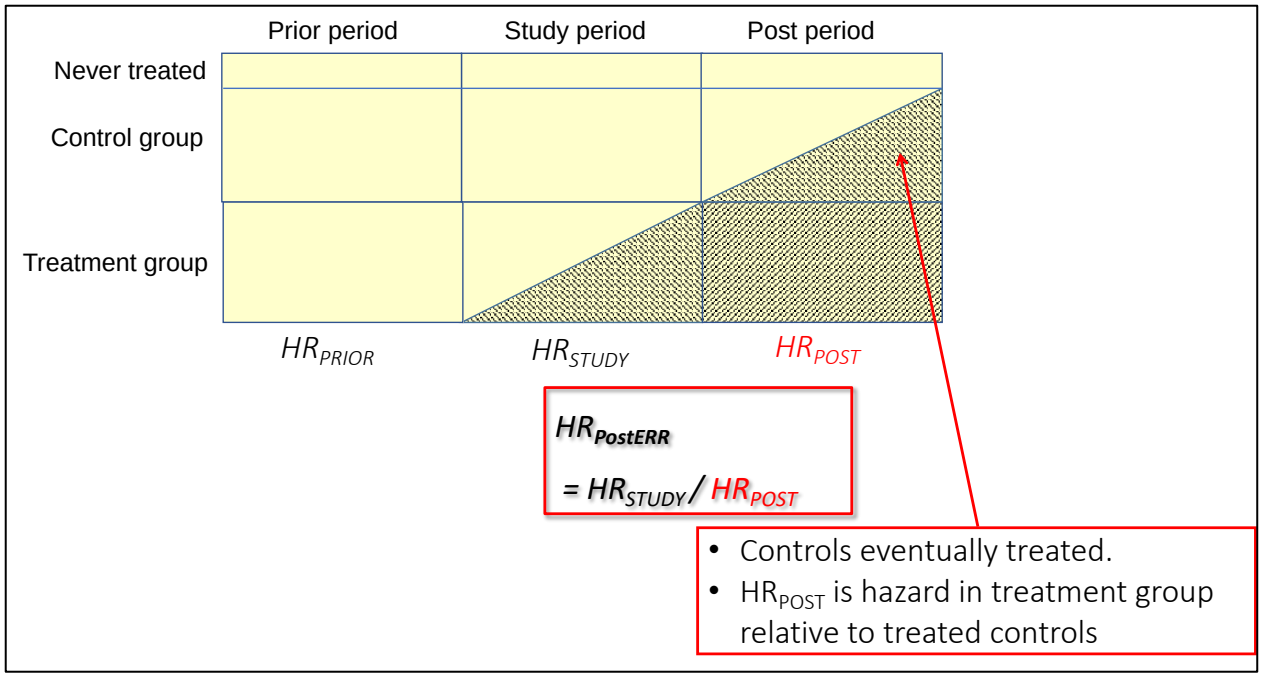


*Figure 40:* Diagram illustrating PostERR method relative to the prior period. As per the cumulative vaccination graph in figure 1, 100% coverage of treatment is rarely achieved at the population level, so there is likely to be a rump of patients, who will never be treated.

This PostERR method, as I have dubbed it, may at a very superficial level seem similar to Tannen and Yu's post-treated event rate ratio (PTERR) method. However, their PTERR method was developed to deal with confounding by unrepeatable events, like death. Rather than using a common all-treated period as with the PostERR method, the PTERR method exploits differences in the as-treated and intention-to-treat (ITT) periods for individual patients in order to adjust for confounding. This, of course, implies applicability to treatments, where non-compliance is possible in the first place, and would preclude investigations into vaccinations and cures. The PTERR method also relies on matching exposed to unexposed patients for comparison of the as-treated and ITT periods, and so might not quite be considered a true quasi-experimental method.

As alternative adjustment for confounding, the PostERR method may offer some advantages instead of, or as a complement to the PERR method:

1. It offers an alternative measure of confounding bias for a sub-group within the PERR cohort, which eventually reaches an exposed state. A different confounding bias may manifest in the post period compared to the prior, so some simulation work will be necessary to investigate this. High-dimensional

adjustment methods, such as IPTW, should be used as a check and verification of the bias apportioned to observed confounding, as followed in this PhD. With the transition of the unexposed into an exposed state, this approach has parallels with the active-comparator new-user design, and similarly could be considered a form of restriction to patients, who will all eventually reach an exposed state.

2. It is clear from the (pneumococcal) vaccination rates over time, there remains a proportion of patients, who are never vaccinated. This would also likely be true for other cures or long-lasting treatments. In the older population, some of these may be terminally ill, or may have died during that year of data collection. However, the never-treated patients may in some way be different from the remaining cohort, depending on the cut-off for determining those, who would never be treated. By considering the PostERR cohort, this should provide another useful insight into the nature of the confounders biasing the study-period results.
3. For the PostERR cohort, inference could potentially be made about not only the size of confounding bias, but also its trend over time by comparing the all-exposed post period with the unexposed prior period. An assumption of the PERR method, applicable to the PostERR method too, is that of time-invariant confounding bias. However, information from the prior and post periods not only offer insight into changes in the bias over time, but potentially an adjustment for it too.

In the first instance, the PostERR method could be demonstrated on the pneumococcal vaccine investigation in chapter five. However, further simulation work would be needed to clarify the above ideas 1-3, as well as the assumptions of PERR, particularly in the presence of confounding by indication.

### 7.4.3 Future research: Methodological development

#### *7.4.3.1 Validation of PERR and Pairwise applied to interactions*

As far as could be ascertained from the published literature, the vaccination studies were the first to apply the PERR and Pairwise methods to moderating effects. In clinical trials, the focus is usually on estimating the efficacy of an intervention as a main effect, with clinically interesting sub-groups subordinated as secondary

analysis. With routinely collected data, there is little justification for being restricted by this paradigm. Also, testing interactions are essential for investigating moderating effects, and for performing global tests of significance to avoid multiplicity across categories of patient characteristics divided into subgroups. With the size and depth of information available for patient histories in EHRs, QE methods, such as the PERR and Pairwise methods could be utilised in identifying responders to treatment and adverse events contributing to the development of personalised medicine. In this PhD project, the pneumococcal vaccination found effectiveness increased with age, contrary to some counter arguments based on the observed phenomenon of immunosenescence. This was verified across a variety of approaches including a non-QE, weighted regression, all reaching the same conclusion about the moderating effect of age. Clearly this has potential to impact on the decisions taken by policy makers. However, this and future work could benefit from further validation of the analysis in which interaction effects were adjusted using the PERR and Pairwise methods in the presence of confounding. It is known from previous simulation work that the Pairwise method, and to a lesser extent the PERR method, are biased in the presence of an interaction between the exposure and any hidden confounder. However, little is known about the performance of the methods when analysing an interaction in the presence of unmeasured confounders. In Figure 41, if M is the moderator, with which exposure X interacts to cause observed Y, then X and Y may be confounded by unmeasured confounders, U, or that U may be aliased with the interaction effect through M, or may directly moderate the effect of X on Y as an unmeasured moderator. All scenarios are also possible. Hence, a simulation is necessary to test the performance of the methods under these conditions. This could be presented in the context of the vaccination studies in the older population, in which age is often a key effect moderator, as well as confounder.

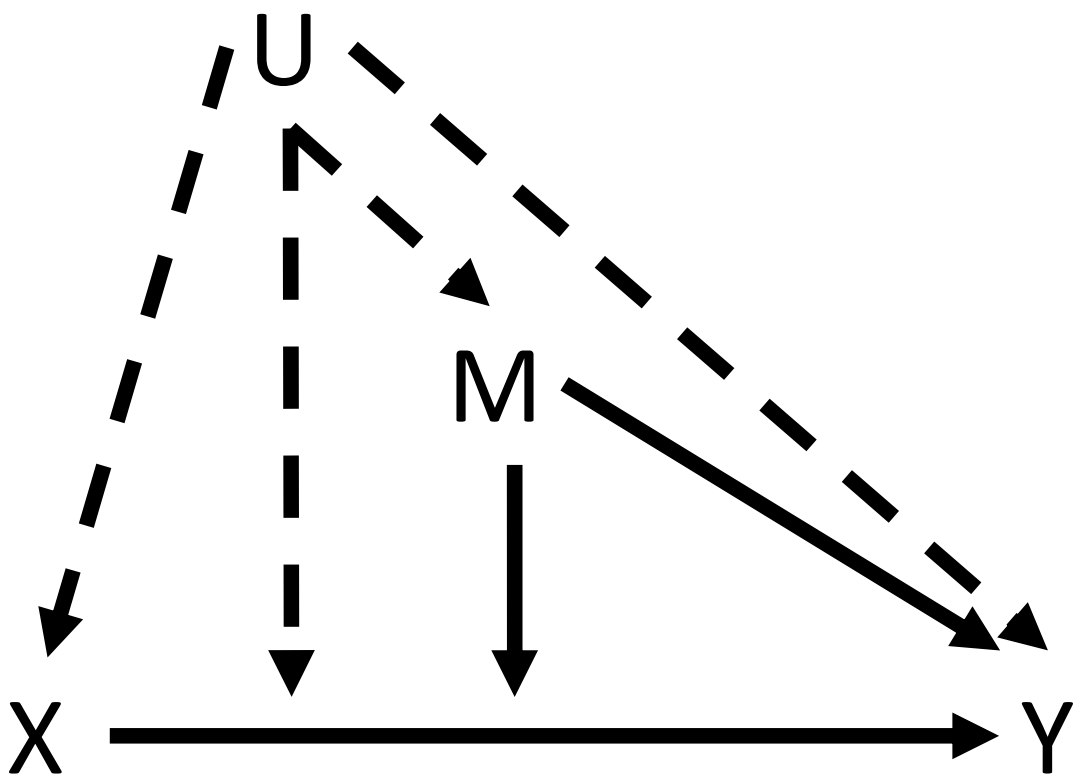


*Figure 41:* Causal diagram showing moderating effect of M on the effect of X on outcome, Y. U is an unmeasured variable and potential confounder and/or mediator. The dashed lines show possible causal pathways to be simulated in the study of the performance of the PERR and Pairwise methods in analysing an interaction.

### *7.4.3.2 Sensitivity to timing of prior period*

As the prior period is key to the adjustment in the PERR and Pairwise methods, then further work might be undertaken to explore the sensitivity of results to varying timings and durations of the prior period. The period immediately prior to the study period was chosen in the vaccination studies as this seemed to afford the best assurance that the confounding bias would be representative of that in the study period. However, the timing of the prior period could be chosen to create a gap between the two periods. This could provide further insight into the stability of the confounding bias over time as well as potentially test the sensitivity to indication bias (effect of prior event on exposure). For interventions like vaccination, some caution would be necessary about inferences into the presence, or otherwise, of indication, as vaccines tend to be routinely offered during a specific season in the year (usually

October to December). In this case, events occurring in the previous season may still indicate vaccination, which would nevertheless be administered in the following season, and so a lag of a whole year may be necessary to entirely separate indication effects. However, as discussed in paragraph four of section 4.3 in chapter four, indication is unlikely in theory to operate completely independently of confounding variables, and so simulation work may be required to determine the point, at which the effect of indication "overtakes" the imbalance in prognostic factors. For other interventions independent of season, the lag between periods could be varied over months rather than in years. However, to study the stability of confounders over time, then the lags may need to be specified in years as well as months to study stability of long term and seasonal unmeasured confounders, respectively.

### 7.4.4 Future research: Best practice guidelines

Presentation of the study on pneumococcal vaccination led to an invitation from the International Society of Pharmacoepidemiology (ISPE) to review the pharmacoepidemiological extension to the pre-existing guidelines for the reporting of studies conducted using observational routinely-collected health data (RECORD), thus creating the RECORD-PE guidelines. This has now been published on the ISPE website [343]. The RECORD guidelines were themselves an extension of the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines adapted for RCD such as EHRs. The RECORD guidelines presented timely recommendations for maintaining good practice in reporting studies based on RCD, justified by the growing interest in using this data in health research. While it was entirely logical to build upon these existing guidelines, there is an argument for starting afresh, and use the opportunity to review and renew existing items, rather than append to them. One of my primary concerns was the separation of items 4 (study design), 9 (bias) and 12 (statistical methods). While EHRs of RCD may form the basis for many ecological studies (e.g: vaccination coverage; changes in prescribing patterns over time), there is an increasing focus on how to tackle the problem of confounding bias and improve the robustness of causal inference from evidence based on EHRs. In this way, study design and statistical methods have been shown (section 4.4 of chapter four) to be two integral strategies for dealing with confounding bias. As confounding bias is arguably the greatest threat to causal

inference from observational data, then this perhaps warrants greater focus in any guidelines for using RCD. This certainly figured prominently in the objective the PE extensions were intended to serve. However, the danger of separating the two complementary strategies in items 4 and 12 for dealing with the property of bias listed in 9, is that that focus becomes too diffuse. While it is helpful to break down the problem into numbered steps in order of reporting, there is an argument for reorganising the guidelines and starting again, perhaps tailoring these for inferential studies, given that adjustment for bias will determine both simultaneously the study design and analytical method used.

An important contribution in the PE-extension to the guidelines was the expansion of the section on study design to focus on how this could be used to mitigate for confounding bias. The explanation included, as an example, the relatively recent development of the active-comparator new-user design, which, as already described, controls for confounding by restricting the study to a comparison between two drugs for the same disease, and therefore, to patients with the same prognosis. However, the examples also included interrupted time series, which it could be argued is an analytical method rather than design, relying as it does on longitudinally collected data, but not necessarily within individuals. The explanation also offered risk-minimisation interventions as an example of study, which was perhaps better suited as an example of a study's purpose rather than its design.

It was further noted that in the explanation for the "Methods (bias)" section that this seemed to overlap with the following explanation for item 12 on statistical methods by including comments on adjustment for observed confounders through propensity score methods. The explanation for item 12a also exclusively focussed on case-only designs, which may erroneously give the impression of being the only acceptable design and method. Future collaboration may be needed for further iterative improvements on the guidelines. However, it is clear that if a contribution is to be made to RCD methodologies, then further work is first required to develop best practice in the application of the PERR and Pairwise methods, building on the work presented in this PhD project. Besides publication of the studies on vaccination, a review based on chapter 4 will also published, which will include the recommendations for best practice in applying the PERR and Pairwise methods,

along with examples where the conditions for the application of the methods may be met.

## 7.5 Conclusions

Through the investigation of vaccine effectiveness in older adults using electronic health records, my work for this PhD project has demonstrated how evidence from routinely collected data can been used to complement the results from clinical trials to build a more complex understanding of a potential clinical effects in real-world settings. I demonstrated how this involved quasi-experimental methods to adjust for confounding bias, and in the application to the vaccination effectiveness studies, I developed a strategy for understanding the source and size of the bias. In these studies, I was able to compare the performance of two related and recently developed methods for adjusting for unmeasured confounding. I also reported on the limitations of these methods, the PERR and Pairwise, and their performance in their application to vaccine effectiveness studies. I also demonstrated the usefulness of these methods in studying populations, such as older adults, that may be under-represented in trials.

My contribution to research into these specific methods is that by drawing together existing research, I was able to report on the relative performance of the PERR and Pairwise methods, and offer guidance for their future application. In a wider context of adjusting for unmeasured confounding, I have also provided published guidance for using quasi-experimental methods in studies using routinely-collected data, and how the available longitudinal information can be leverage to adjust for unmeasured confounding. For future research, I have identified avenues for improving the applicability of the PERR approach, and potentially its robustness against time-varying confounding.

My research also had a clinical impact. Through the novel application of the PERR and Pairwise methods to estimating effect modification, I presented evidence for an increase in effectiveness of the pneumococcal vaccine with age in older adults, during the implementation of the UK vaccination programme. In my study of the influenza vaccine in the same population, I presented evidence for the protective effect of the influenza vaccine against myocardial infarction, and how this is likely to be mediated through its protective effect against influenza. I also highlighted the challenges in applying these methods to vaccination studies, but identifed another

area for its application, studying the drug interaction between statins and the influenza vaccine and presented my protocol for this, based on my pilot work.

# Appendix A – methodological review search terms

1. ("prior event" and ratio).ti,ab.

2. "paired cox model".ti,ab.

3. 1 or 2

4. instrumental variables.ti,ab.

5. instrumental variable analysis/

6. propensity score calibration.ti,ab.

7. regression discontinuity design.ti,ab.

8. "difference in differences".ti,ab.

9. (difference adj1 differences).ti,ab.

10. "ratio of ratios".ti,ab.

11. (ratio adj1 ratios).ti,ab.

12. interrupted time series.ti,ab.

13. segmented regression.ti,ab.

14. (sensitivity analysis/ or sensitivity analysis.ti,ab.) and ((unmeasured or residual or hidden) and (confounding or confounder*)).ti,ab.

15. or/4-14

16. ((unmeasured or residual or hidden or unobserved or omitted) and (confounding or confounder*)).ti,ab.

17. confounding variable/

18. covariates.ti,ab.

19. bias.ti,ab.

20. selection bias/

21. 16 or 17 or 18 or 19 or 20

22. observational study/

23. (observation* adj (stud* or data)).ti,ab.

24. ((before adj after) and (study or studies)).ti,ab.

25. (nonrandomi?ed or non randomi?ed).ti,ab.

26. case crossover.ti,ab.

27. case control.ti,ab.

28. case control study/

29. cohort study.ti,ab.

30. (quasi experiment* or quasiexperiment*).ti,ab.

31. quasi-experimental study/

32. cross sectional study.ti,ab.

33. cross-sectional study/

34. simulation.ti,ab.

35. case time control.ti,ab.

36. ("before and after" and (study or studies)).ti,ab.

37. or/22-36

38. 16 and 19 and 37

39. 3 or 15

40. 39 and 37 and 21

41. 38 or 40

42. 21 or 37

43. 39 and 42

# Appendix B – table of studies included in the methodological review

*Table 25: Table of included studies denoting QE method used and type of instrument, if applicable, where: IVA = instrumental variable analysis; RD = regression discontinuity; DiD = difference-in-differences; DiDiD = difference-in-difference-in-differences; PSC = propensity score calibration; PERR = prior event rate ratio*

| Author | Title | Year | QE method | If IVA, IV type |
|---|---|---|---|---|
| Bryson, W. C.; McConnell, J.; Krothuis, T.; McCarty, D. | Extended-release naltrexone for alcohol dependence: persistence and healthcare costs and utilization | 2011 | DiD | |
| Cheng, L.; Liu, H.; Zhang, Y.; Shen, K.; Zeng, Y. | The impact of health insurance on health outcomes and spending of the elderly: Evidence from china's new cooperative medical scheme | 2015 | DiD | |
| Gebel, M.; Vosemer, J. | The impact of employment transitions on health in Germany. A difference-in-differences propensity score matching approach | 2014 | DiD | |

| | | | |
|---|---|---|---|
| Goetzel, R. Z.; Roemer, E. C.; Pei, X.; Short, M. E.; Tabrizi, M. J.; Wilson, M. G.; Dejoy, D. M.; Craun, B. A.; Tully, K. J.; White, J. M.; Baase, C. M. | Second-year results of an obesity prevention program at the dow chemical company | 2010 | DiD |
| Higgins, S.; Chawla, R.; Colombo, C.; Snyder, R.; Nigam, S. | Medical homes and cost and utilization among high-risk patients | 2014 | DiD |
| Kausto, J.; Viikari-Juntura, E.; Virta, L. J.; Gould, R.; Koskinen, A.; Solovieva, S. | Effectiveness of new legislation on partial sickness benefit on work participation: a quasi-experiment in Finland | 2014 | DiD |
| Kelly, Y.; Kelly, J.; Sacker, A. | Changes in bedtime schedules and behavioral difficulties in 7 year old children | 2013 | DiD |
| Lin, W. C.; Chien, H. L.; Willis, G.; O'Connell, E.; Rennie, K. S.; Bottella, H. M.; Ferris, T. G. | The effect of a telephone-based health coaching disease management program on medicaid members with chronic conditions | 2012 | DiD |

| | | | |
|---|---|---|---|
| Lyon, S. M.; Wunsch, H.; Asch, D. A.; Carr, B. G.; Kahn, J. M.; Cooke, C. R. | Use of intensive care services and associated hospital mortality after massachusetts healthcare reform | 2014 | DiD |
| Menon, J.; Paulet, M.; Thomas, Iii J. | Wellness coaching and health-related quality of life: A case-control difference-in-differences analysis | 2012 | DiD |
| Moran, J. R.; Short, P. F.; Hollenbeak, C. S. | Long-term employment effects of surviving cancer | 2011 | DiD |
| Osborne, N. H.; Nicholas, L. H.; Ryan, A. M.; Thumma, J. R.; Dimick, J. B. | Association of hospital participation in a quality reporting program with surgical outcomes and expenditures for medicare beneficiaries | 2015 | DiD |
| Reid, R. O.; Ashwood, J. S.; Friedberg, M. W.; Weber, E. S.; Setodji, C. M.; Mehrotra, A. | Retail clinic visits and receipt of primary care | 2013 | DiD |
| Sadhu, A. R.; Ang, A. C.; Ingram-Drake, L. A.; Martinez, D. S.; Hsueh, W. A.; Ettner, S. L. | Economic benefits of intensive insulin therapy in critically Ill patients: The targeted insulin therapy to improve hospital outcomes (TRIUMPH) project | 2008 | DiD |

| | | | |
|---|---|---|---|
| Sarkar, U.; Lyles, C. R.; Parker, M. M.; Allen, J.; Nguyen, R.; Moffet, H. H.; Schillinger, D.; Karter, A. J. | Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system | 2014 | DiD |
| Watt, C.; Abuya, T.; Warren, C. E.; Obare, F.; Kanya, L.; Bellows, B. | Can reproductive health voucher programs improve quality of postnatal care? A quasi-experimental evaluation of Kenya ' s Safe Motherhood voucher scheme | 2015 | DiD |
| De Preux, L. B. | Anticipatory ex ante moral hazard and the effect of medicare on prevention | 2011 | DiD; DiDiD |
| Rajaram, R.; Chung, J. W.; Jones, A. T.; Cohen, M. E.; Dahlke, A. R.; Ko, C. Y.; Tarpley, J. L.; Lewis, F. R.; Hoyt, D. B.; Bilimoria, K. Y. | Association of the 2011 ACGME resident duty hour reform with general surgery patient outcomes and with resident examination performance | 2014 | DiD; DiDiD |
| Domino, M. E.; Norton, E. C.; Morrissey, J. P.; Thakur, N. | Cost shifting to jails after a change to managed mental health care | 2004 | DiD; Fixed effects |

| | | | | |
|---|---|---|---|---|
| Hodgkin, D.; Parks Thomas, C.; Simoni-Wastila, L.; Ritter, G. A.; Lee, S. | The effect of a three-tier formulary on antidepressant utilization and expenditures | 2008 | Fixed effects | |
| Li, J.; Hurley, J.; DeCicca, P.; Buckley, G. | Physician response to pay-for-performance: evidence from a natural experiment | 2014 | DiD pooled OLS; DiD (Fixed effects); DiD + differential trends | |
| Yoon, J.; Bernell, S. L. | The role of adverse physical health events on the utilization of mental health services | 2013 | DiD & Fixed Effects | |
| Fortney, J. C.; Steffick, D. E.; Burgess Jr, J. F.; Maciejewski, M. L.; Petersen, L. A. | Are primary care services a substitute or complement for specialty and inpatient services? | 2005 | IVA applied to DiD | Geographic |
| Hay, J.; Jhaveri, M.; Tangirala, M.; Kaliner, M. | Cost and resource utilization comparisons of second-generation antihistamines vs. montelukast for allergic rhinitis treatment | 2009 | IVA applied to Fixed effects | Historical |
| Chung, S.; Domino, M. E.; Stearns, S. C. | The effect of retirement on weight | 2009 | Fixed Effects; IVA applied to Fixed effects | Lagged |

| | | | | |
|---|---|---|---|---|
| Wagner, T. H.; Jimison, H. B. | Computerized health information and the demand for medical care | 2003 | IVA applied to Fixed effects | Other |
| Kawatkar, A. A.; Hay, J. W.; Stohl, W.; Nichol, M. B. | Incremental expenditure of biologic disease modifying antirheumatic treatment using instrumental variables in panel data | 2013 | Dynamic panel model (IV-GMM) | Lagged |
| Piernas, C.; Ng, S. W.; Mendez, M. A.; Gordon-Larsen, P.; Popkin, B. M. | A dynamic panel model of the associations of sweetened beverage purchases with dietary quality and food-purchasing patterns | 2015 | Dynamic panel model (IV-GMM) | Lagged |
| Lei, X.; Lin, W. | The new cooperative medical scheme in rural China: Does more coverage mean more service and better health? | 2009 | Fixed effects; IVA; DiD | Geographic |
| Lin, M. J.; Liu, J. T. | Do lower birth weight babies have lower grades? Twin fixed effect and instrumental variable method evidence from Taiwan | 2009 | Fixed effects; IVA | Geographic |
| Schmittdiel, J. A.; Karter, A. J.; Dyer, W.; Parker, M.; Uratsu, C.; Chan, J.; Duru, O. K. | The comparative effectiveness of mail order pharmacy use vs. local pharmacy use on LDL-C control in new statin users | 2011 | DiD; IVA | Other |

| | | | | |
|---|---|---|---|---|
| Basu, A. | Estimating Decision-Relevant Comparative Effects Using Instrumental Variables | 2011 | IVA | Geographic |
| Beck, C. A.; Penrod, J.; Gyorkos, T. W.; Shapiro, S.; Pilote, L. | Does Aggressive Care Following Acute Myocardial Infarction Reduce Mortality? Analysis with Instrumental Variables to Compare Effectiveness in Canadian and United States Patient Populations | 2003 | IVA | Geographic |
| Chen, L. F.; Chen, H. P.; Huang, Y. S.; Huang, K. Y.; Chou, P.; Lee, C. C. | Pneumococcal Pneumonia and the Risk of Stroke: A Population-Based Follow-Up Study | 2012 | IVA | Geographic |
| Edwards, S. T.; Prentice, J. C.; Simon, S. R.; Pizer, S. D. | Home-Based Primary Care and the risk of ambulatory care-sensitive condition hospitalization among older veterans with diabetes mellitus | 2014 | IVA | Geographic |
| Frances, C. D.; Shlipak, M. G.; Noguchi, H.; Heidenreich, P. A.; McClellan, M. | Does physician specialty affect the survival of elderly patients with myocardial infarction? | 2000 | IVA | Geographic |
| Goldman, D. P.; Bao, Y. | Effective HIV treatment and the employment of HIV+ adults | 2004 | IVA | Geographic |

| | | | | |
|---|---|---|---|---|
| Gowrisankaran, G.; Town, R. J. | Estimating the quality of care in hospitals using instrumental variables | 1999 | IVA | Geographic |
| Hirth, R. A.; Grabowski, D. C.; Feng, Z.; Rahman, M.; Mor, V. | Effect of nursing home ownership on hospitalization of long-stay residents: An instrumental variables approach | 2014 | IVA | Geographic |
| Kahn, J. M.; Werner, R. M.; David, G.; Ten Have, T. R.; Benson, N. M.; Asch, D. A. | Effectiveness of long-term acute care hospitalization in elderly patients with chronic critical illness | 2013 | IVA | Geographic |
| Linden, A.; Adams, J. L. | Evaluating disease management programme effectiveness: An introduction to instrumental variables | 2006 | IVA | Geographic |
| Norton, E. C.; Lindrooth, R. C.; Ennett, S. T. | Controlling for the endogeneity of peer substance use on adolescent alcohol and tobacco use | 1998 | IVA | Geographic |
| Pilote, L.; Beck, C. A.; Eisenberg, M. J.; Humphries, K.; Joseph, L.; Penrod, J. R.; Tu, J. V. | Comparing invasive and noninvasive management strategies for acute myocardial infarction using administrative databases | 2008 | IVA | Geographic |

| | | | | |
|---|---|---|---|---|
| Pracht, E. E.; Tepas, Iii J. J.; Celso, B. G.; Langland-Orban, B.; Flint, L. | Survival advantage associated with treatment of injury at designated trauma centers: A bivariate probit model with instrumental variables | 2007 | IVA | Geographic |
| Slade, E. P.; McCarthy, J. F.; Valenstein, M.; Visnic, S.; Dixon, L. B. | Cost savings from assertive community treatment services in an era of declining psychiatric inpatient use | 2013 | IVA | Geographic |
| Tsai, A. C.; Votruba, M.; Bridges, J. F. P.; Cebul, R. D. | Overcoming bias in estimating the volume-outcome relationship | 2006 | IVA | Geographic |
| Wehby, G. L.; Ullrich, F.; Xie, Y. | Very low birth weight hospital volume and mortality: An instrumental variables approach | 2012 | IVA | Geographic |
| Hadley, J.; Polsky, D.; Mandelblatt, J. S.; Mitchell, J. M.; Weeks, J. C.; Wang, Q.; Hwang, Y. T. | An exploratory instrumental variable analysis of the outcomes of localized breast cancer treatments in a medicare population | 2003 | IVA | Geographic + Historical + Time |
| O'Malley, A. J.; Frank, R. G.; Normand, S. L. T. | Estimating cost-offsets of new medications: Use of new antipsychotics and mental health costs for schizophrenia | 2011 | IVA | Geographic + Time |

| | | | | |
|---|---|---|---|---|
| Abrahamowicz, M.; Beauchamp, M. E.; Ionescu-Ittu, R.; Delaney, J. A. C.; Pilote, L. | Reducing the variance of the prescribing preference-based instrumental variable estimates of the treatment effect | 2011 | IVA | Historical |
| An, J.; Nichol, M. B. | Multiple medication adherence and its effect on clinical outcomes among patients with comorbid type 2 diabetes and hypertension | 2013 | IVA | Historical |
| Bekelman, J. E.; Mitra, N.; Handorf, E. A.; Uzzo, R. G.; Hahn, S. A.; Polsky, D.; Armstrong, K. | Effectiveness of androgen-deprivation therapy and radiotherapy for older men with locally advanced prostate cancer | 2015 | IVA | Historical |
| Bhowmik, D.; Aparasu, R. R.; Rajan, S. S.; Sherer, J. T.; Ochoa-Perez, M.; Chen, H. | Risk of manic switch associated with antidepressant therapy in pediatric bipolar depression | 2014 | IVA | Historical |
| Brooks, J. M.; Tang, Y.; Chapman, C. G.; Cook, E. A.; Chrischilles, E. A. | What is the effect of area size when using local area practice style as an instrument? | 2013 | IVA | Historical |
| Chuang, C. M.; Chou, Y. J.; Yen, M. S.; Chao, K. C.; Twu, N. F.; Wu, H. H.; Wen, K. C.; Chen, Y. J.; Wang, P. H.; Lai, C. R.; Chou, P. | The role of secondary cytoreductive surgery in patients with recurrent epithelial ovarian, tubal, and peritoneal cancers: A comparative effectiveness analysis | 2012 | IVA | Historical |

| | | | | |
|---|---|---|---|---|
| De Ridder, A.; De Graeve, D. | Can we account for selection bias? A comparison between bare metal and drug-eluting stents | 2011 | IVA | Historical |
| Fang, G.; Brooks, J. M.; Chrischilles, E. A. | Comparison of instrumental variable analysis using a new instrument with risk adjustment methods to reduce confounding by indication | 2012 | IVA | Historical |
| Figueroa, R.; Harman, J.; Engberg, J. | Use of Claims Data to Examine the Impact of Length of Inpatient Psychiatric Stay on Readmission Rate | 2004 | IVA | Historical |
| Huesch, M. D. | External adjustment sensitivity analysis for unmeasured confounding: An application to coronary stent outcomes, Pennsylvania 2004-2008 | 2013 | IVA | Historical |
| Huybrechts, K. F.; Brookhart, M. A.; Rothman, K. J.; Silliman, R. A.; Gerhard, T.; Crystal, S.; Schneeweiss, S. | Comparison of different approaches to confounding adjustment in a study on the association of antipsychotic medication with mortality in older nursing home patients | 2011 | IVA | Historical |

| | | | | |
|---|---|---|---|---|
| Ionescu-Ittu, R. | Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables | 2012 | IVA | Historical |
| Kivimaki, M.; Vahtera, J.; Kawachi, I.; Ferrie, J. E.; Oksanen, T.; Joensuu, M.; Pentti, J.; Salo, P.; Elovainio, M.; Virtanen, M. | Psychosocial work environment as a risk factor for absence with a psychiatric diagnosis: An instrumental-variables analysis | 2010 | IVA | Historical |
| Kramer, A.; Jager, K. J.; Fogarty, D. G.; Ravani, P.; Finne, P.; Perez-Panades, J.; Prutz, K. G.; Arias, M.; Heaf, J. G.; Wanner, C.; Stel, V. S. | Association between pre-transplant dialysis modality and patient and graft survival after kidney transplantation | 2012 | IVA | Historical |
| Kuo, Y. F.; Montie, J. E.; Shahinian, V. B. | Reducing bias in the assessment of treatment effectiveness: Androgen deprivation therapy for prostate cancer | 2012 | IVA | Historical |
| Lakdawalla, D. N.; Mascarenhas, M.; Jena, A. B.; Vanderpuye-Orgle, J.; Lavallee, C.; Linthicum, M. T.; Snider, J. T. | Impact of oral nutrition supplements on hospital outcomes in pediatric patients | 2014 | IVA | Historical |

| | | | | |
|---|---|---|---|---|
| MacKenzie, T. A.; Tosteson, T. D.; Morden, N. E.; Stukel, T. A.; O'Malley, A. J. | Using instrumental variables to estimate a Cox's proportional hazards regression subject to additive confounding | 2014 | IVA | Historical |
| Margolis, D. J.; Gupta, J.; Hoffstad, O.; Papdopoulos, M.; Glick, H. A.; Thom, S. R.; Mitra, N. | Lack of effectiveness of hyperbaric oxygen therapy for the treatment of diabetic foot ulcer and the prevention of amputation a cohort study | 2013 | IVA | Historical |
| Parmar, A. D.; Sheffield, K. M.; Han, Y.; Vargas, G. M.; Guturu, P.; Kuo, Y. F.; Goodwin, J. S.; Riall, T. S. | Evaluating comparative effectiveness with observational data: Endoscopic ultrasound and survival in pancreatic cancer | 2013 | IVA | Historical |
| Pisoni, R. L.; Arrington, C. J.; Albert, J. M.; Ethier, J.; Kimata, N.; Krishnan, M.; Rayner, H. C.; Saito, A.; Sands, J. J.; Saran, R.; Gillespie, B.; Wolfe, R. A.; Port, F. K. | Facility Hemodialysis Vascular Access Use and Mortality in Countries Participating in DOPPS: An Instrumental Variable Analysis | 2009 | IVA | Historical |
| Prentice, J. C.; Conlin, P. R.; Gellad, W. F.; Edelman, D.; Lee, T. A.; Pizer, S. D. | Capitalizing on prescribing pattern variation to compare medications for type 2 diabetes | 2014 | IVA | Historical |

| | | | | |
|---|---|---|---|---|
| Rassen, J. A.; Brookhart, M. A.; Glynn, R. J.; Mittleman, M. A.; Schneeweiss, S. | Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance | 2009 | IVA | Historical |
| Rosenthal, M. B.; Li, Z.; Robertson, A. D.; Milstein, A. | Impact of financial incentives for prenatal care on birth outcomes and spending | 2009 | IVA | Historical |
| Sheffield, K. M.; Riall, T. S.; Han, Y.; Kuo, Y. F.; Townsend, C. M., Jr.; Goodwin, J. S. | Association between cholecystectomy with vs without intraoperative cholangiography and risk of common duct injury | 2013 | IVA | Historical |
| Steingrub, J. S.; Lagu, T.; Rothberg, M. B.; Nathanson, B. H.; Raghunathan, K.; Lindenauer, P. K. | Treatment with neuromuscular blocking agents and the risk of in-hospital mortality among mechanically ventilated patients with severe sepsis | 2014 | IVA | Historical |
| Stukel, Thérèse A; Fisher, Elliott S; Wennberg, David E; Alter, David A; Gottlieb, Daniel J; Vermeulen, Marian J | Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. | 2007 | IVA | Historical |

| | | | | |
|---|---|---|---|---|
| Tagami, T.; Matsui, H.; Horiguchi, H.; Fushimi, K.; Yasunaga, H. | Antithrombin and mortality in severe pneumonia patients with sepsis-associated disseminated intravascular coagulation: An observational nationwide study | 2014 | IVA | Historical |
| VanDyke, R. D.; McPhail, G. L.; Huang, B.; Fenchel, M. C.; Amin, R. S.; Carle, A. C.; Chini, B. A.; Seid, M. | Inhaled tobramycin effectively reduces FEV1 decline in cystic fibrosis an instrumental variables analysis | 2013 | IVA | Historical |
| Wong, K.; Campitelli, M. A.; Stukel, T. A.; Kwong, J. C. | Estimating influenza vaccine effectiveness in community-dwelling elderly patients using the instrumental variable analysis method | 2012 | IVA | Historical |
| Chen, H.; Mehta, S.; Aparasu, R.; Patel, A.; Ochoa-Perez, M. | Comparative effectiveness of monotherapy with mood stabilizers versus second generation (atypical) antipsychotics for the treatment of bipolar disorder in children and adolescents | 2014 | IVA | Historical + Time |

| | | | | |
|---|---|---|---|---|
| Newman, T. B.; Vittinghoff, E.; McCulloch, C. E. | Efficacy of phototherapy for newborns with hyperbilirubinemia: a cautionary example of an instrumental variable analysis | 2012 | IVA | Historical + Time |
| Ahern, T. P.; Pedersen, L.; Svaerke, C.; Rothman, K. J.; Sorensen, H. T.; Lash, T. L. | The association between vitamin K antagonist therapy and site-specific cancer incidence estimated by using heart valve replacement as an instrumental variable | 2011 | IVA | Lagged |
| Cai, B.; Hennessy, S.; Flory, J. H.; Sha, D.;Ten Have, T. R.; Small, D. S. | Simulation study of instrumental variable approaches with an application to a study of the antidiabetic effect of bezafibrate | 2012 | IVA | Lagged |
| O'Malley, A. J. | Instrumental variable specifications and assumptions for longitudinal analysis of mental health cost offsets | 2012 | IVA | Lagged |
| Cawley, J.; Meyerhoefer, C. | The medical care costs of obesity: An instrumental variables approach | 2012 | IVA | Other |

| | | | | |
|---|---|---|---|---|
| Groenwold, R. H.; Hak, E.; Klungel, O. H.; Hoes, A. W. | Instrumental variables in influenza vaccination studies: mission impossible?! | 2010 | IVA | Other |
| Kim, D.; Leigh, J. P. | Estimating the effects of wages on obesity | 2010 | IVA | Other |
| Pirracchio, R.; Sprung, C.; Payen, D.; Chevret, S. | Benefits of ICU admission in critically ill patients: whether instrumental variable methods or propensity scores should be used | 2011 | IVA | Other |
| Selden, T. M.; Hudson, J. L. | Access to care and utilization among children: Estimating the effects of public and private coverage | 2006 | IVA | Other |
| Slade, E. P.; Wissow, L. S.; Davis, M.; Abrams, M. T.; Dixon, L. B. | Medicaid lapses and low-income young adults' receipt of outpatient mental health care after an inpatient stay | 2014 | IVA | Other |
| Hay, J. W.; Lawler, E.; Yucel, K.; Guo, A.; Balzer, T.; Gaziano, J. M.; Scranton, R. E. | Cost impact of diagnostic imaging for lower extremity peripheral vascular occlusive disease | 2009 | IVA | PScore (historical EHRs) |

| | | | | |
|---|---|---|---|---|
| Guo, J.; Konetzka, R. T.; Manning, W. G. | The causal effects of home care use on institutional long-term care utilization and expenditures | 2015 | IVA | Randomisation |
| Federspiel, J. J.; Stearns, S. C.; Sheridan, B. C.; Kuritzky, J. J.; D'Arcy, L. P.; Crespin, D. J.; Carey, T. S.; Rossi, J. S. | Evaluating the effectiveness of a rapidly adopted cardiovascular technology with administrative data: The case of drug-eluting stents for acute coronary syndromes | 2012 | IVA | Time |
| Goyal, N.; Zubizarreta, J. R.; Small, D. S.; Lorch, S. A. | Length of stay and readmission among late preterm infants: An instrumental variable approach | 2013 | IVA | Time |
| Hollingsworth, J. M.; Norton, E. C.; Kaufman, S. R.; Smith, R. M.; Wolf Jr, J. S.; Hollenbeck, B. K. | Medical expulsive therapy versus early endoscopic stone removal for acute renal colic: An instrumental variable analysis | 2013 | IVA | Time |
| Johnston, K. M.; Gustafson, P.; Levy, A. R.; Grootendorst, P. | Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research | 2008 | IVA | Time |

| | | | | |
|---|---|---|---|---|
| O'Donnell, H. C.; Colman, G.; Trachtman, R. A.; Velazco, N.; Racine, A. D. | Impact of newborn follow-up visit timing on subsequent ED visits and hospital readmissions: AN instrumental variable analysis | 2014 | IVA | Time |
| Zeliadt, S. B.; Loggers, E. T.; Slatore, C. G.; Au, D. H.; Hebert, P. L.; Klein, G. J.; Kessler, L. G.; Backhus, L. M. | Preoperative PET and the reduction of unnecessary surgery among newly diagnosed lung cancer patients in a community setting | 2014 | IVA | Time |
| Brunner, E. J.; Kivimaki, M.; Witte, D. R.; Lawlor, D. A.; Davey Smith, G.; Cooper, J. A.; Miller, M.; Lowe, G. D.; Rumley, A.; Casas, J. P.; Shah, T.; Humphries, S. E.; Hingorani, A. D.; Marmot, M. G.; Timpson, N. J.; Kumari, M. | Inflammation, insulin resistance, and diabetes--Mendelian randomization using CRP haplotypes points upstream | 2008 | IVA (Mendelian) | Mendelian |
| Burgess, S.; Thompson, S. G. | Avoiding bias from weak instruments in mendelian randomization studies | 2011 | IVA (Mendelian) | Mendelian |

| | | | | |
|---|---|---|---|---|
| Haring, R.; Teumer, A.; Volker, U.; Dorr, M.; Nauck, M.; Biffar, R.; Volzke, H.; Baumeister, S. E.; Wallaschofski, H. | Mendelian randomization suggests non-causal associations of testosterone with cardiometabolic risk factors and mortality | 2013 | IVA (Mendelian) | Mendelian |
| Jokela, M.; Elovainio, M.; Keltikangas-Jarvinen, L.; Batty, G. D.; Hintsanen, M.; Seppala, I.; Kahonen, M.; Viikari, J. S.; Raitakari, O. T.; Lehtimaki, T.; Kivimaki, M. | Body mass index and depressive symptoms: Instrumental-variables regression with genetic risk score | 2012 | IVA (Mendelian) | Mendelian |
| Kivimaki, M.; Magnussen, C. G.; Juonala, M.; Kahonen, M.; Kettunen, J.; Loo, B. M.; Lehtimaki, T.; Viikari, J.; Raitakari, O. T. | Conventional and Mendelian randomization analyses suggest no association between lipoprotein(a) and early atherosclerosis: The Young Finns Study | 2011 | IVA (Mendelian) | Mendelian |

| | | | | |
|---|---|---|---|---|
| Laschkolnig, A.; Kollerits, B.; Lamina, C.; Meisinger, C.; Rantner, B.; Stadler, M.; Peters, A.; Koenig, W.; Stockl, A.; Dahnhardt, D.; Boger, C. A.; Kramer, B. K.; Fraedrich, G.; Strauch, K.; Kronenberg, F. | Lipoprotein (a) concentrations, apolipoprotein (a) phenotypes, and peripheral arterial disease in three independent cohorts | 2014 | IVA (Mendelian) | Mendelian |
| Lawlor, D. A.; Harbord, R. M.; Timpson, N. J.; Lowe, G. D.; Rumley, A.; Gaunt, T. R.; Baker, I.; Yarnell, J. W.; Kivimaki, M.; Kumari, M.; Norman, P. E.; Jamrozik, K.; Hankey, G. J.; Almeida, O. P.; Flicker, L.; Warrington, N.; Marmot, M. G.; Ben-Shlomo, Y.; Palmer, L. J.; Day, I. N.; Ebrahim, S.; Smith, G. D. | The association of C-reactive protein and CRP genotype with coronary heart disease: findings from five studies with 4,610 cases amongst 18,637 participants | 2008 | IVA (Mendelian) | Mendelian |

| | | | | |
|---|---|---|---|---|
| Leong, A.; Rehman, W.; Dastani, Z.; Greenwood, C.; Timpson, N.; Langsetmo, L.; Berger, C.; Fu, L.; Wong, B. Y. L.; Malik, S.; Malik, R.; Hanley, D. A.; Cole, D. E. C.; Goltzman, D.; Richards, J. B. | The Causal Effect of Vitamin D Binding Protein (DBP) Levels on Calcemic and Cardiometabolic Diseases: A Mendelian Randomization Study | 2014 | IVA (Mendelian) | Mendelian |

| | | | | |
|---|---|---|---|---|
| Nimptsch, K.; Aleksandrova, K.; Boeing, H.; Janke, J.; Lee, Y. A.; Jenab, M.; Bueno-De-Mesquita, H. B.; Jansen, E. H. J. M.; Tsilidis, K. K.; Trichopoulou, A.; Weiderpass, E.; Wu, C.; Overvad, K.; Tjonneland, A.; Boutron-Ruault, M. C.; Dossus, L.; Racine, A.; Kaaks, R.; Canzian, F.; Lagiou, P.; Trichopoulos, D.; Palli, D.; Agnoli, C.; Tumino, R.; Vineis, P.; Panico, S.; Johansson, A.; Van Guelpen, B.; Khaw, K. T.; Wareham, N.; Peeters, P. H.; Quiros, J. R.; Garcia, A. V.; Molina-Montes, E.; Dorronsoro, M.; Chirlaque, M. D.; Gurrea, A. B.; Key, T. J.; Duarte-Salles, T.; Stepien, M.; Gunter, M. J.; Riboli, E.; Pischon, T. | Association of CRP genetic variants with blood concentrations of C-reactive protein and colorectal cancer risk | 2015 | IVA (Mendelian) | Mendelian |

| | | | | |
|---|---|---|---|---|
| Palmer, T. M.; Sterne, J. A. C.; Harbord, R. M.; Lawlor, D. A.; Sheehan, N. A.; Meng, S.; Granell, R.; Smith, G. D.; Didelez, V. | Instrumental variable estimation of causal risk ratios and causal odds ratios in mendelian randomization analyses | 2011 | IVA (Mendelian) | Mendelian |
| Wehby, G. L.; Scholder, Sv | Genetic instrumental variable studies of effects of prenatal risk factors | 2013 | IVA (Mendelian) | Mendelian |
| Richardson, D. B.; Laurier, D.; Schubauer-Berigan, M. K.; Tchetgen, E. T.; Cole, S. R. | Assessment and indirect adjustment for confounding by smoking in cohort studies using relative hazards models | 2014 | Negative Control Outcome | |
| Brophy, S.; Jones, K. H.; Rahman, M. A.; Zhou, S. M.; John, A.; Atkinson, M. D.; Francis, N.; Lyons, R. A.; Dunstan, F. | Incidence of campylobacter and salmonella infections following first prescription for PPI: A cohort study using routine data | 2013 | PERR | |
| Tannen, R. L. | Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: Comparison of database and randomised controlled trial findings | 2009 | PERR | |

| | | | |
|---|---|---|---|
| Tannen, R. L.; Weiner, M. G.; Xie, D. | Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: Further empiric validation of the 'prior event rate ratio' to adjust for unmeasured confounding by indication | 2008 | PERR |
| Tannen, R.; Xie, D.; Wang, X.; Yu, M.; Weiner, M. G. | A new "Comparative Effectiveness" assessment strategy using the THIN database: Comparison of the cardiac complications of pioglitazone and rosiglitazone | 2013 | PERR |
| Uddin, M. J.; Groenwold, R. H. H.; Van Staa, T. P.; De Boer, A.; Belitser, S. V.; Hoes, A. W.; Roes, K. C. B.; Klungel, O. H. | Performance of prior event rate ratio adjustment method in pharmacoepidemiology: A simulation study | 2015 | PERR |
| Lee, W. C. | Detecting and correcting the bias of unmeasured factors using perturbation analysis: a data-mining approach | 2014 | Perturbation analysis |
| Lunt, M.; Glynn, R. J.; Rothman, K. J.; Avorn, J.; Sturmer, T. | Propensity score calibration in the absence of surrogacy | 2012 | PSC |

| | | | |
|---|---|---|---|
| Sturmer, T. | Performance of propensity score calibration - A simulation study | 2007 | PSC |
| Stürmer, Til, Schneeweiss, Sebastian, Avorn, Jerry; | Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration | 2005 | PSC |
| Albouy, V.; Lequien, L. | Does compulsory education lower mortality? | 2009 | RD |
| Swaminathan, S.; Mor, V.; Mehrotra, R.; Trivedi, A. N. | Effect of medicare dialysis payment reform on use of erythropoiesis stimulating agents | 2015 | RD |
| Zuckerman, I. H.; Lee, E.; Wutoh, A. K.; Xue, Z.; Stuart, B. | Application of regression-discontinuity analysis in pharmaceutical health services research | 2006 | RD |

# Appendix C – ISAC protocol

Below is the final accepted version of protocol submitted to ISAC for approval. Highlighted text in blue indicates changes that have been made at the request of the initial ISAC review. Yellow highlights indicates text that was added since the previous version.

ISAC APPLICATION FORM
PROTOCOLS FOR RESEARCH USING THE CLINICAL PRACTICE RESEARCH DATALINK (CPRD)

| | | |
|---|---|---|
| ISAC use only:<br><br>Protocol Number<br><br>Date submitted | <br><br>.....<br><br>..... | **IMPORTANT**<br><br>**If you have any queries, please contact ISAC Secretariat:**<br>ISAC@cprd.com |

| |
|---|
| 1. Study Title<br>**Influenza and Pneumococcal Vaccination Effectiveness and an Investigation into Association with Cardiovascular Outcomes in the Elderly** |
| 2. Principal Investigator (full name, job title, organisation & e-mail address for correspondence regarding this protocol)<br>**David Melzer**<br><br>**Professor of Epidemiology and Public Health**<br><br>**Epidemiology and Public Health Group**<br><br>**University of Exeter**<br><br>**Medical School**<br><br>**Email: D.Melzer@exeter.ac.uk** |
| 3. Affiliation (full address)<br>**University of Exeter**<br><br>**Medical School**<br><br>**Barrack Road, Exeter, EX2 5DW**<br><br>**United Kingdom** |
| 4. Protocol's Author (if different from the principal investigator)<br>**PI plus Alessandro Ble, Adam Streeter and William Henley** |
| 5. List of all investigators/collaborators (*please list the names, affiliations and e-mail addresses* of all collaborators*, other than the principal investigator) |

**Alessandro Ble**

**Research Fellow in Epidemiology and Public Health**

**Epidemiology and Public Health Group**

**University of Exeter**

**Medical School**

**Email: A.Ble@exeter.ac.uk**

**Kirsty Bowman**

**PhD student in Epidemiology and Public Health**

**Epidemiology and Public Health Group**

**University of Exeter**

**Medical School**

**Email: khb202@exeter.ac.uk**

**William Henley**

**Professor of Medical Statistics**

**Health Statistics Group**

**University of Exeter**

**Medical School**

**Email: W.E.Henley@exeter.ac.uk**

**Jane Masoli**

**NIHR Academic Clinical Fellow & Specialist Registrar in Geriatric Medicine,**

**Epidemiology and Public Health Group & Healthcare for Older People, RD&E Hospital, Exeter**

**University of Exeter**

**Medical School**

**Email: J.Masoli@exeter.ac.uk**

**Ruben Mujica Mota,**

**Senior Lecturer in Economics**

**University of Exeter**

**Medical School**

**Email: E.Mujica-Mota@exeter.ac.uk**

**Suzanne Richards,**

**Senior Lecturer in Primary Care,**

**Primary Care Research Group**

**University of Exeter**

**Medical School**

**Email: S.H.Richards@exeter.ac.uk**

**Adam Streeter**

**PhD student in Medical Statistics**

**Health Statistics Group**

**University of Exeter**

**Medical School**

**Email: A.J.Streeter@exeter.ac.uk**

**Jose M. Valderas,**

**Professor of Health Services and Policy Research & Academic General Practitioner**

**Primary Care Research Group**

**University of Exeter**

**Medical School**

**Email: J.M.Valderas@exeter.ac.uk**

**Lauren Rogers**

**Research fellow in Medical Statistics**

**Health Statistics Group**

**University of Exeter**

**Medical School**

**L.R.Rodgers@exeter.ac.uk**

**Please note that your ISAC application form and protocol **must** be copied to all e-mail addresses listed above at the time of submission of your application to the ISAC mailbox. Failure to do so will result in delays in the processing of your application.*

6. Type of Institution (please tick one box below)

| | | | | | |
|---|---|---|---|---|---|
| Academia | ☒ | Research Service Provider | ☐ | Pharmaceutical Industry | ☐ |
| NHS | ☐ | Government Departments | ☐ | Others | ☐ |

7. Financial Sponsor of study

| | | | | | |
|---|---|---|---|---|---|
| Pharmaceutical Industry *(please specify)* | ☐ | | | Academia *(please specify)* | ☐ |
| Government / NHS *(please specify)* | ☒ | NIHR | | None | ☐ |
| Other *(please specify)* | ☐ | | | | |

8. Data source *(please tick one box below)*

| | | | |
|---|---|---|---|
| Sponsor has on-line access | ☐ | Purchase of ad hoc dataset | ☐ |
| Commissioned study | ☐ | | |
| Other | ☒ *(please specify)* | amended use of 14-135R database | |

9. Has this protocol been peer reviewed by another Committee?

Yes* ☐ No ☒ Note: sub-study within NIHR School for Public Health Ageing Well research programme – programme proposal peer reviewed and approved

** Please state in your protocol the name of the reviewing Committee(s) and provide an outline of the review process and outcome.*

10. Type of Study *(please tick all the relevant boxes which apply)*

Adverse Drug Reaction/Drug Safety ☐ Drug Use ☒ Disease Epidemiology ☒

Drug Effectiveness ☒ Pharmacoeconomic ☐ Other ☐

11. This study is intended for:

Publication in peer reviewed journals ☒ Presentation at scientific conference ☒

Presentation at company/institutional meetings ☐ Other

12. Does this protocol also seek access to data held under the CPRD Data Linkage Scheme?

Yes ☒ No ☐

13. If you are seeking access to data held under the CPRD Data Linkage Scheme*, please select the source(s) of linked data being requested.

☒ Hospital Episode Statistics ☐ Cancer Registry Data**

☐ MINAP ☒ ONS Mortality Data

☒ Index of Multiple Deprivation/ Townsend Score

☐ Mother Baby Link ☐ Other: (please specify)

** As part of the ISAC review of linkages, the protocol may be shared - in confidence - with a representative of the requested linked data set(s) and summary details may be shared - in confidence - with the Confidentiality Advisory Group of the Health Research Authority.*

***Please note that applicants seeking access to cancer registry data must provide consent for publication of their study title and study institution on the UK Cancer Registry website. Please contact the CPRD Research Team on +44 (20) 3080 6383 or email kc@cprd.com to discuss this requirement further.*

14. If you are seeking access to data held under the CPRD Data Linkage Scheme, have you already discussed your request with a member of the Research team?

Yes ☒ No* ☐

**Please contact the CPRD Research Team on +44 (20) 3080 6383 or email kc@cprd.com to discuss your requirements before submitting your application.*

Please list below the name of the person/s at the CPRD with whom you have discussed your request.

**Kendal Chidwick**

15. If you are seeking access to data held under the CPRD Data Linkage Scheme, please provide the following information:

The number of linked datasets requested: 4

A synopsis of the purpose(s) for which the linkages are required:

**We request the CPRD standard data (for the linked subsample), plus HES data to cover hospital event outcomes. We also request ONS mortality data as these seem to improve ascertainment of death over GP recorded date of death data, at least in the oldest old (from our previous work). Permission for MINAP data is also requested, if possible, as the study is focussed on CVD outcomes**

Is linkage to a local dataset with <1 million patients being requested?

Yes* ☐ No ☒

** If yes, please provide further details:*

16. If you have requested linked data sets, please indicate whether the Principal Investigator or any of the collaborators listed in response to question 5 above, have access to any of the linked datasets in a patient identifiable form, or associated with a patient index.

Yes* ☐ No ☒

** If yes, please provide further details:*

17. Does this protocol involve requesting any additional information from GPs?

Yes* ☐ No ☒

** Please indicate what will be required:*

Completion of questionnaires by the GP[ψ] Yes ☐ No ☐

Provision of anonymised records (e.g. hospital discharge summaries) Yes ☐ No ☐

Other (please describe)

*[ψ] Any questionnaire for completion by GPs or other health care professional must be approved by ISAC before circulation for completion.*

18. Does this protocol describe a purely observational study using CPRD data (this may include the review of anonymised free text)?

Yes* ☒ No** ☐

** Yes: If you will be using data obtained from the CPRD Group, this study does not require separate ethics approval from an NHS Research Ethics Committee.*

*** No: You may need to seek separate ethics approval from an NHS Research Ethics Committee for this study. The ISAC will provide advice on whether this may be needed.*

19. Does this study involve linking to patient *identifiable* data from other sources?

Yes ☐ No ☒

20. Does this study require contact with patients in order for them to complete a questionnaire?

Yes ☐ No ☒

*N.B. Any questionnaire for completion by patients must be approved by ISAC before circulation for completion.*

21. Does this study require contact with patients in order to collect a sample?

Yes* ☐ No ☒

** Please state what will be collected*

22. Experience/expertise available

Please complete the following questions to indicate the experience/expertise available within the team of researchers actively involved in the proposed research, including analysis of data and interpretation of results

| | Previous GPRD/CPRD Studies | Publications using GPRD/CPRD data |
|---|---|---|
| None | ☐ | ☐ |
| 1-3 | ☒ | ☒ |
| > 3 | ☐ | ☐ |

| | Yes | No |
|---|---|---|
| Is statistical expertise available within the research team? | ☒ | ☐ |

*If yes, please outline level of experience*

Co-I William Henley is Professor of Medical Statistics with extensive experience in methods, including accounting for missing values. Co-I Adam Streeter is a PhD student in Medical Statistics. PI David Melzer has got more than 15 years of experience in analysing data from large studies on ageing and from genomic array datasets.

| | | |
|---|---|---|
| Is experience of handling large data sets (>1 million records) available within the research team? | ☒ | ☐ |

*If yes, please outline level of experience*

PI David Melzer and Co-I William Hanley have 2-year experience in analysing a 50,000 patient extract of the CPRD data. CI Alessandro Ble and CI Adam Streeter have a 1-year experience in working with the CPRD database. DM has got more than 15-years' experience in analysing very large genomic array datasets, including UK Biobank.

| | | |
|---|---|---|
| Is UK primary care experience available within the research team? | ☒ | ☐ |

*If yes, please outline level of experience*

Suzanne Richards is Senior Lecturer in Primary Care and Dr Jose Valderas is academic general practitioner and Professor of Health Services and Policy Research at the University of Exeter. They will provide extensive inputs on primary care issues.

23. References relating to your study

Please list up to 3 references (most relevant) relating to your proposed study.

Jefferson T, Di Pietrantonj C, Al-Ansary LA, Ferroni E, Thorning S, Thomas RE. Vaccines for preventing influenza in the elderly. Cochrane database Syst Rev. 2010;(2):CD004876. doi:10.1002/14651858.CD004876.pub3.

Moberley S, Holden J, Tatham DP, Andrews RM. Vaccines for preventing pneumococcal infection in adults. Cochrane database Syst Rev. 2013;1:CD000422. doi:10.1002/14651858.CD000422.pub3.

**PROTOCOL CONTENT CHECKLIST**

In order to help ensure that protocols submitted for review contain adequate information for protocol evaluation, ISAC have produced instructions on the content of protocols for research using CPRD data. These instructions are available on the CPRD website (www.cprd.com/ISAC). All protocols using CPRD data which are submitted for review by ISAC must contain information on the areas detailed in the instructions. IF you do not feel that a specific area required by ISAC is relevant for your protocol, you will need to justify this decision to ISAC.

Applicants must complete the checklist below to confirm that the protocol being submitted includes all the areas required by ISAC, or to provide justification where a required area is not considered to be relevant for a specific protocol. Protocols will not be circulated to ISAC for review until the checklist has been completed by the applicant.

**Please note, your protocol will be returned to you if you do not complete this checklist, or if you answer 'no' and fail to include justification for the omission of any required area.**

| | **Included in protocol?** | | |
|---|---|---|---|
| Required area | Yes | No | If no, reason for omission |
| *Lay Summary (max.200 words)* | ☒ | ☐ | |
| *Background* | ☒ | ☐ | |
| *Objective, specific aims and rationale* | ☒ | ☐ | |
| *Study Type*<br>*Descriptive*<br>*Hypothesis Generating*<br>*Hypothesis Testing* | <br>☐<br>☐<br>☒ | <br>☒<br>☒<br>☐ | <br>Hypothesis testing<br>Hypothesis testing |
| *Study Design* | ☒ | ☐ | |
| *Sample size/power calculation*<br>*(Please provide justification of sample size in the protocol)* | ☒ | ☐ | |
| *Study population*<br>*(including estimate of expected number of* | ☒ | ☐ | |

| | | | |
|---|---|---|---|
| *relevant patients in the CPRD)* | | | |
| *Selection of comparison group(s) or controls* | ☒ | ☐ | |
| *Exposures, outcomes and covariates*<br>*Exposures are clearly described*<br>*Outcomes are clearly described* | ☒<br>☒<br>☒ | ☐<br>☐<br>☐ | |
| *Use of linked data*<br>*(if applicable)* | ☒ | ☐ | |
| *Data/ Statistical Analysis Plan*<br>*There is plan for addressing confounding*<br>*There is a plan for addressing missing data* | ☒<br>☒<br>☐ | ☐<br>☐<br>☒ | please see |
| *Patient/ user group involvement* † | ☒ | ☐ | |
| *Limitations of the study design, data sources and analytic methods* | ☒ | ☐ | |
| *Plans for disseminating and communicating study results* | ☒ | ☐ | |

† ***It is expected that many studies will benefit from the involvement of patient or user groups in their planning and refinement, and/or in the interpretation of the results and plans for further work. This is particularly, but not exclusively true of studies with interests in the impact on quality of life. Please indicate whether or not you intend to engage patients in any of the ways mentioned above.***

***Voluntary registration of ISAC approved studies:***
*Epidemiological studies are increasingly being included in registries of research around the world, including those primarily set up for clinical trials. To increase awareness amongst researchers of ongoing research, ISAC encourages voluntary registration of epidemiological research conducted using MHRA databases. This will not replace information on ISAC approved protocols that may be published in its summary minutes or annual report. It is for the applicant to determine the most appropriate registry for their study. Please inform the ISAC secretariat that you have registered a protocol and provide the location.*

## Lay summary

Adults aged over 65y are presently targeted in the UK for vaccination against influenza and pneumococcal infection. Yet recent studies have suggested a possible additional protective effect against major cardiovascular events. Evidence from randomized clinical trials to support the policy of vaccination and the effect on cardiovascular outcomes is sparse due to the ethical and practical difficulties of recruiting a representative sample of elderly subjects, many of who have multiple co-morbidities. Observational data can offer an information-rich addition to trial evidence. However analysis of such data is subject to confounding bias, with studies into vaccine effectiveness variously failing to control for the effects of unmeasured confounding. Therefore evaluation of vaccine effectiveness needs to be robust against the effects of unmeasured confounders. Accordingly, the analysis in this study will be validated against other analytical methods and the results compared with estimates synthesised from existing evidence and current epidemiological data. Pneumococcal pneumonia is a common complication of influenza, and both present as respiratory disease, so the effects of both vaccines should be considered together. This project is part of the NIHR School for Public Health Research 'Ageing Well' research programme and extends work on risk factors and prevention of cardiovascular disease in different clinically relevant groups of older patients.

## Objectives, Specific Aims and Rationale

The objective of this project is to estimate the effects of vaccination against Influenza and Pneumococcal infection in the elderly (age 65 and over), on:

a) rates of lower respiratory infection, measured through hospital diagnoses and primary care prescribing

b) major cardiovascular events, to reappraise previous claims of major protective effects.

To provide valid results we will employ and compare two statistical methods, which aim to deal with bias and confounding, namely propensity scoring and the prior event rate ratio (PERR). In addition to producing estimates of effect, we aim also to understand the nature and source of bias present in this observational study context.

The elderly population is diverse, and we therefore plan to estimate vaccination effects in younger and older groups separately in men and women and in those with different burdens of co-morbidity

## Background

The overall objective of the NIHR National School for Public Health Research 'ageing well' research programme is to help improve the evidence base for prevention and health promotion in later life. This CPRD application is focused on General Practitioner (GP) delivered influenza and pneumococcal vaccination, which is claimed to have a role in cardiovascular prevention in older people.

Vaccination against influenza viruses and the streptococcus pneumoniae (pneumococcus) bacterium is currently recommended in adults aged over 65y and is intended to tackle age-related incidents of lower respiratory tract infection, of which the most acute form is pneumonia. Estimated to be the fifth leading cause of mortality in adults aged over 65y with prevalence highest in the very old [344,345], pneumonia is increasingly a cause for hospitalisation in the elderly [346]. Conversely deaths from influenza are difficult to discern without laboratory confirmation, but while morality rates due to influenza per se are estimated to be low compared to those from pneumonia [345], the US Centers for Disease Control and Prevention recommend vaccination against influenza to reduce the risk of complications, such as pneumonia, in the at-risk populations. In the USA, it has been estimated that influenza and its complications are responsible on average for 186 000 excess hospitalisations[347] and 132.5 deaths per 100 000 person-years[348].

Reliable estimates of the benefits of vaccination are important for establishing informed policies regarding resource allocation and identifying the need for new vaccines and prevention strategies. However, there is currently a lack of robust evidence for effectiveness of influenza and pneumococcal vaccines in older populations:

a) A meta-analysis of studies evaluating the effectiveness of vaccination against pneumococcal infection in adults [280] found immunisation to be effective against culture-confirmed invasive pneumococcal disease, but provided inconclusive results for the respiratory disease of pneumonia, by far the most common presentation of infection. The effect of routine vaccination to prevent all-cause pneumonia and mortality among the elderly, therefore, remains unresolved.

b) Observational studies evaluating the effectiveness of the influenza vaccine have so far largely proved to be unsatisfactory, failing to adequately control for indication bias [349], while in the only major randomised controlled trial in the elderly using laboratory-confirmed outcomes, those above 70y of age were under-represented [350].

c) In addition to the intended effect of vaccination, a recent meta-analysis of RCT evidence on influenza vaccination[312] (n=6735 patients, mean age 67 years) concluded that vaccination was associated with a risk reduction of about 35% from major adverse cardiovascular events, but recommended further investigation on a larger scale. Furthermore, there is some evidence suggesting that pneumococcal vaccination may also be associated with a reduced risk of myocardial infarction[351] although this has been less consistently found than for influenza vaccination[352,353].

Older patients included in clinical trials often tend to be selected according to criteria that are more narrowly representative of the general population. In this project we aim to estimate 'real world' effects of pneumococcal and influenza vaccination in the whole elderly population. However, the older population is very diverse, with for example disease-free 67 year olds having little in common with 85 year olds on medication for three or more major conditions. We therefore seek to estimate vaccination effects in defined major sub-groups of the older population, based on age, gender and co-morbidity.

## Study type

The study will principally be a hypothesis test of the effectiveness of the pneumococcal and influenza vaccines against the respiratory diseases caused by those pathogens targeted for immunisation and major cardiovascular outcomes. Based on experience and evidence from previous studies, confounding bias presents a major challenge in the analysis of

observational data. Therefore the analysis plan includes the comparison of the effects of measured confounding against adjustment made for residual confounding.

## Study design

This study is designed as a parallel matched cohorts study. This study involves statistical analysis of existing data from GP and linked hospital records for patients aged 65 and over. We plan to apply two statistical methods analyses to the data in order to address potential confounding, comparing patients who received each vaccination with similar patients who did not. We will apply propensity scoring, which seeks to match patients on measures that predict the outcomes of interest. We will also use the PERR approach (see statistics section), which aims to use the rates of the outcome of interest (e.g. rates for lower respiratory infections) before vaccination to correct for prior differences between the case group who got vaccinations and a control group who didn't.

## Data/Statistical analysis

The purpose of this study is to evaluate the true effect of vaccination against influenza and pneumococcal infection, and to verify recently reported secondary effects on cardiovascular disease. As the data are observational, the best method will be sought to adjust for any confounding bias, that is likely to be unmeasured and otherwise retained in the residual of any analysis. In doing so, we aim to understand the nature of this bias, and compare the contribution from measured and unmeasured sources. This will require comparing the different approaches to the analysis, as described below.

The data will be modelled as survival times from the vaccination date (or equivalent date for the controls) to the outcome of interest. Vaccine effectiveness will initially be estimated using Cox's regression – the fundamental model to which adjustment for confounding bias will be applied.

*Propensity scoring*

An important step in the analysis will be to assess the contribution of known confounders where they are provided in the data. Vaccination and the outcomes will individually be tested for association with covariates and factors identified as potential confounders through multivariate modelling. Significant variables will be modelled in a best-fit model of vaccination status. This will be used to generate predictions, or propensity scores[11]. The literature on how to apply the propensity score methodology is extensive [212,294,354,355] and shall not be repeated here, but the treatment-prediction model generating the scores needs to be specified correctly in order to adjust for confounding. Proper adjustment for confounding bias is of course contingent on correct specification of all variables associated with both treatment and outcome. The treatment effect with adjustment for known confounders through the propensity score can then be estimated using Cox's regression. Here we are interested in the marginal effect, rather than any conditional effects, so bias due to non-collapsibility of the non-linear regression model will be avoided.

*The prior event rate ratio*

Since routinely collected data such as that supplied by CPRD is unlikely to have collected information on all confounders relating to any particular study, further methods are required to adjust for confounding bias, which would otherwise remain in the residual of analysis.

The Prior Event Rate Ratio (PERR) method does not require identification of individual confounders and can be applied to the hazard ratio of vaccine effectiveness from the Cox's regression. This quasi-experimental analytic method requires knowledge of event rates in the vaccinated and control patients during a vaccination-free period prior to the period under study. The assumption is made that the ratio of outcome events in the vaccinated and control groups during the prior period will reflect the combined effect of all identified and unidentified confounders related to that outcome. Having estimated the hazard ratio of treatment between the two groups in both prior and study periods through Cox's regression, the hazard ratio from the treatment-free prior period is used to adjust the hazard ratio of treatment in the study period. Applying PERR to time-to-event data from the

Clinical Practice Research Datalink in order to estimate effectiveness of treatment for hypertension yielded convincing results that control for confounding bias can be achieved[188] study by Yu et al [246] provided further support for the validity of the method, suggesting that the method is robust to deviations from key assumptions.

The alternative formulation of the PERR method will also be applied. The PERR-ALT method differs slightly in that the prior period data is used to adjust the study period within each treatment group, before hazard ratio for the treatment is derived. The pairwise adjustment of PERR-ALT within each treatment arm could be viewed as analogous to that of the self-controlled case series method. While the authors of PERR suggest that PERR-ALT is robust to interactions between unmeasured confounders and time intervals in the presence of relatively large treatment effects, the PERR method itself is computationally more stable when the events are rare. The purpose of applying both PERR and PERR-ALT will be to arrive at the best unbiased estimate, and in doing so, demonstrate the level of bias that may be inherent from the effects of hidden covariates in the PERR treatment of the nonlinear Cox model [225].

The hazard ratio in the treatment-free prior period will be used to gauge the level of unmeasured confounding and the results from the PERR-adjusted and PERR-ALT models compared to those achieved through propensity score adjusted models (recommended by Yu et al) and basic models adjusted for gender and age. The interaction of both vaccine statuses will be tested for all outcomes.

Results from the analysis of cardiovascular outcomes will be reported for clinically important sub-groups defined by age-band, gender and beyond a certain level of existing co-morbidities (i.e: in top 30% of count) : see study population patient subgroups.

Stata v13 and R will be used in the analysis.

Missing data would likely comprise longitudinal measurement such as blood pressure or weight. Such data would be best imputed deploying the two-fold fully conditional specification (FCS) algorithm [356], available in Stata as the *twofold*, with attention paid to the specification in the imputed models of survival time, the outcome of subsequent substantive models fitted to the imputed data [357]. An assumption of missing-at-random will be required for the covariates. The data will be explored to investigate the plausibility of

being missing-at-random. For key inferences, sensitivity analysis performed to test departures from this assumption. The results from imputed data will be compared to those from complete-case analysis to assess the level of bias corrected through imputation and whether the magnitude and direction of bias is consistent with prior expectations.

**Sensitivity analysis**

It can be argued that first and subsequent myocardial infarctions can differ with respect to underlying mechanism and associated treatments.

We therefore plan the following sensitivity analyses:

1. We will add terms relating to first or subsequent MI and revascularisation (receipt and procedure, from HES data) as a confounder in the propensity score models and the PERR proportional hazards models

2. We will conduct a subgroup analyses excluding those with second and subsequent MIs

3. if possible in the full dataset, we will examine those with at least one prior MI before the period of interst, and then model the effect of flu vaccination on subsequent MI incidence

## Sample size

Vaccination effectiveness will be estimated from Cox regression models. Based on estimates from the 2012 edition of the Coronary Heart Disease Statistics, from the British Heart Foundation, women aged over 84y, a clinically interesting subgroup, experienced of 139 incidences of acute myocardial infarction (a primary outcome for ischemic heart disease) per 100 000 person-years. As the meta-analysis into studies of influenza vaccination on cardiovascular disease found the effect could reduce the risk by about 35%, the power to detect a hazard ratio of 0.65 is sought at a power of 0.8 and significance level of 5%. Using the Schoenfeld approximation, this would require a sample size of 121 710 patients. If the power to detect a hazard ratio of, say, 0.70 were sought, this would increase to 177 550 patients.

The sample size required for the preliminary investigation of intended effect of the vaccinations on respiratory conditions is rather less demanding. An estimate of incidence for community-acquired pneumonia among adults aged over 65y put the rate at 7.99 per 1000 person-years from CPRD data. Therefore the probability of survival times ending in such an outcome over a three-year study period of the pneumococcal vaccine would be approximately 0.024. Results from a meta analysis of studies into the pneumococcal vaccine suggested that the risk of pneumococcal pneumonia could be reduced by approximately 16%. Therefore to detect a hazard ratio of 0.85 at a power of 0.8 and significance level of 5% would require 49 530 eligible patients.

The Melzer group is arranging an institutional CPRD annual academic licence supported by the NIHR School for Public Health Research Ageing Well programme funding. This will provide online access to the full CPRD-Gold database of patients aged 65 and over, with linked HES and mortalty data, estimated at 1.6 million patients aged 65 and over. This will provide the most robust estimates possible, and given the above minimum sample size requirements the full database should afford more than an acceptable level of power.

Sample size calculations were performed in Stata v.13.

### Data linkage

While suspected pneumonia cases may be treated with antibiotics prescribed within primary care, hospitalisation is a common consequence of pneumonia in the elderly. For this reason access to Hospital Episode Statistics (HES) is integral to the study of vaccines against respiratory diseases. Analysis will be conducted using antibiotic outcomes in all available patients in the cohort. Additionally we will analyse hospitalisations with a pneumonia diagnosis, combining these with the primary care antibiotic outcomes into a composite outcome among the HES-linked patients only.

Secondarily we will examine admissions with a cardiovascular disease code.

Given the advancing age of the cohort, linkage to ONS-sourced death dates is required to ensure a reliable record of death.

### Study population

The policy of routine immunisation against influenza and pneumococcus was originally extended to adults aged over 65y in 2000 and 2003, respectively. Immunisation against influenza is recommended every year, primarily due to the evolution of the influenza viruses, while that against pneumococcus is officially regarded as conferring lifetime immunity. Monitoring by the Health Protection Agency/ Public Health England shows that coverage for pneumococcal vaccination has risen from a baseline of below 30% before 2003 to the current level of 69.1% in England, while that for influenza has witnessed an increase from 46% in 1999 to current levels between 71% and 75%.

We would therefore use an extracted dataset of adults who have reached the age of 65y since the year 1997. The timing is especially critical for analysis of the pneumococcal vaccine, whose effect is accepted as lifelong. Given the post-millennial increase in vaccine coverage, data from the period preceding the policy change is needed to deliver estimates from a vaccine-free period for a sufficient number of yet-to-be-vaccinated elderly individuals. Analysis of this treatment-free period prior to intervention is necessary to evaluate the level of confounding bias between two treatment groups (see figure 1 in Appendix B). Integral to the PERR method, explained in the Data/Statistical Methods section, the treatment estimate from this prior period is used to account for unmeasured confounding. Adults aged at least 65y at the start of the study would subsequently be recruited from the extracted data to the cohort for the study period of interest. Further inclusion criteria for each cohort are that the subjects should be continually registered at their GP practice for the duration of the study and prior periods.

In making a quasi-experimental comparison with a prior period, the principal method for our analysis is only applicable to recurrent events in that these should be possible in both the study and prior periods. Death, particularly from the diseases subjected to vaccination, is of interest and shall be investigated with regard to how much imbalance might be created between the analyses of the prior and study periods. Since individuals need to have survived into the study period, sampling from the extracted data should of course exclude deaths in the prior period.

Another consequence of the lifelong effect of the pneumococcal vaccination is that over time the increasing coverage since 2003 effectively reduces the pool of non-vaccinees from which to recruit controls. Using the pilot data, we calculated that a study period of three

years would optimise the size of the recruited sample without reducing the ratio of controls to vaccinees to below unity. The responses in the three-year treatment-free prior period, preceding the study period would control for bias, subject to the assumptions required for the PERR method, outlined in the Statistical Analysis section. To study the effectiveness of the influenza vaccine, the study and prior period will be restricted to one year, covering each influenza season.

The sample size calculations dictate that a cohort size of at least 50 000 adults aged over 65y would be required for the necessary preliminary investigation into vaccine effectiveness against intended respiratory outcomes. However to answer the principle research question looking at effectiveness against cardiovascular events, we would foresee a full extraction to the maximum permissible size of 300K would be justified to allow annual estimates of specific sub-groups to be reported with sufficient power.

## Patient subgroups:

Sub-group analysis will be restricted to a pre-specified hypothesis. Age and gender are commonly considered to be essential variables for sub-group analysis. Consistent with the School for Public Health Ageing Well programme, a fundamental comparison of two age groups per gender will carried out through a simple sub-group analysis of two age bands (65 to 84y and 85+y) for each gender, the null hypothesis being that the response is the same in the “oldest old” as it is for less-elderly patients in either gender.

Another major aim of the programme is to determine the risk from the level of pre-existing co-morbidities at the time of exposure. We plan to compute a simple count of diseases and syndromes and analyse as a sub-group those patients in the top 30% of the count in each gender. The results will be compared with those of all patients in the study within each respective gender group. At the time of writing the conditions and syndromes to be counted are those based on our previous work on diagnostic trends in the oldest old (Melzer et al, Age and Ageing 2014), which are:

*Hypertension, atrial fibrillation, CVD, heart failure, stroke, cancer, chronic kidney disease (stages 3-5) asthma, COD, Dementia, Depression, major mental conditions, epilepsy, diabetes, hypothyroidism, anaemia, osteoarthritis and osteoporosis. Five geriatric syndromes were identified from Read codes for dizziness (including vertigo and syncope), incontinence (urinary and faecal), skin ulcers (including bed sores), falls and fractures.*

It should be stressed that consistent with Bland (1995) that only a count of these conditions will be used as comorbidity index, rather than individual sub-groups and so are not subject to the type 1 errors of multiple testing.

NHS England has supported the development of a similar measure of frailty to be used in electronic clinical records (http://www.hsj.co.uk/resource-centre/supplements/primary-care-supplement-an-index-of-frailty/5065467.article#.U-OFqvldV8E), based on approximately 3000 Read codes, in the TPP system. Publication of the code list and validation studies is expected shortly (Dr A P Clegg, **Clinical Senior Lecturer & Honorary Consultant Geriatrician | Academic Unit of Elderly Care and Rehabilitation | Bradford Institute for Health Research,** private communication). We also plan to apply this approach in a sensitivity analysis, if the validation is convincing.

## Selection of controls

Observation time in the treated group will begin with the date of exposure so that immortal time bias, though commonly associated with cohort studies, will not be an issue for this study. Furthermore the vaccination date will lag by the period required to establish a full immune response following vaccination. Consideration will also be given to the lag between the date, by which symptoms are presented, and the probable time of infection.

The start of the study period, or index date, will depend on the vaccination under study. For influenza vaccine, this will be from the 1st September of each year, to coincide with the start of the season, during which vaccination typically occurs. For the pneumococcal vaccine, there is one single study period beginning on 1st September 2002 for a duration that

encompasses the increase in vaccination coverage arising from the implementation of the immunisation policy.

Patients that remain unexposed to the vaccine throughout the study period will be matched as controls by age, gender and GP practice to the vaccinees from the same period. Observation times in the controls will start from the vaccination dates of the treated patients, with whom they are matched, rather than begin at the start date of the study period. Mapping treatment dates through matched individuals will avoid further bias if the outcome exhibits strong periodicity, for example with seasons, within the follow-up period. Both controls and vaccinees shall have a treatment-free period, necessary for the quasi-experimental adjustment for confounding bias. The duration should be equal to the period over which the patients were recruited for the study and conclude with the start of the study period. In the prior period, each patient's observation period starts at the date relative to the start date of the prior period mapped from the study period relative to the index date.

## Exposures, outcomes and covariates

Exposure

Vaccination status is recorded on the date of vaccination (*eventdate*) and coded (*medcode*) in the Immunisation file (see Appendix).

Outcomes

The outcomes of interest for the effect on major cardiovascular events will be incident cardiovascular disease, comprising incident stroke, myocardial infarction, coronary artery bypass graft and percutaneous transluminal coronary angioplasty, recorded in Hospital HES data. Coding approaches to classifying these are set out in the Appendix and are based on QoF business rules.

Acute lower respiratory disease is the most common presentation of infection by either pneumococcus bacteria or the influenza viruses. In the absence of routine laboratory tests for both conditions, investigators must rely on correct diagnosis and the subsequent accurate recording in the clinical database. Antibiotic prescriptions are commonly prescribed to treat patients presenting with LRTIs and recommended by clinical guidelines

for the more acute form, pneumococcal pneumonia. Both diseases are common complications of infection by the influenza viruses. Where pneumococcal pneumonia is strongly suspected, then cases are frequently hospitalised[290]. Accordingly, hospitalisations for pneumonia have previously been used as an outcome, but the effect on prescribing rates for antibiotics in primary care should also be taken into account. Therefore besides independent evaluations of both outcomes, a composite comprising both outcomes will be analysed. A composite endpoint will be the first occurrence, since the start of the survival time, of either an antibiotic prescription for a respiratory infection in general practice, or a pneumonia related hospitalisation in HES, whichever occurs first.

We will include antibiotics identified as appropriate in the British Thoracic Society guidelines on community acquired respiratory infections. We will identify product codes found in the CPRD *Therapy* file corresponding to the formulations given in the British National Formulary.

The identified cases of antibiotic prescriptions will be further qualified by symptoms for LRTI as described by the corresponding codes (*medcodes*) from the *Clinical* file. Two clinicians will independently identify these. A third clinician will arbitrate any discordance between the two.

Hospitalisation as recorded in the HES files, for HES-linked patients only, will be qualified by the primary reason for episodes as denoted by the ICD10 code (see Annex for codes).

Covariates

Matching for the purpose of mapping observation start times to controls will be based on variables, age, gender and GP practice as recorded in the patient files.

Covariates and factors identified a priori as potential confounders may be included as adjustments to the analysis following the procedure for selection, detailed in the *Data/Statistical Analysis* section below. Besides basic demographic variables of age, gender and socio-economic status, confounders will be identified from the 15 diseases listed under the Quality Outcomes Framework of 2010. Read codes for each disorder have already been obtained through interrogation of the CPRD and HES data using version 18 of the QOF Business Rules, along with a summary of existing useable codes for each disorder (Salisbury 2013). Consensus within the Age UK Project team at the University of Exeter Medical School

decided which QOF disease areas were included and which were not pertinent to a definite diagnosis. The final disease categories were: coronary heart disease, heart failure, stroke and transient ischaemic attack, hypertension, diabetes, chronic obstructive pulmonary disorder (COPD), epilepsy, hypothyroidism, cancer, mental health (includes schizophrenia, bipolar affective disorder and other psychoses), asthma, dementia, depression, chronic kidney disease (CKD), and atrial fibrillation. Smoking history will also be elicited from the records and treated as a potential confounder.

Informed by the work from the "Estimating Cardiovascular Risk in the Elderly" study within the same project team, further confounders will be considered for analysis of cardiovascular events. These will include treatments for existing cardiovascular risk factors: cholesterol-lowering drugs, anti-hypertensive drugs, anti-diabetic drugs, anti-platelets, oral anti-coagulants, beta blockers, calcium antagonists, diuretics, angiotensin-converting enzyme inhibitors, angiotensin II antagonists, nitrates.

Where data permits body mass index, alcohol consumption and total cholesterol / hypercholesterolemia diagnosis including low-density and high-density lipids will also be considered.

For the cardiovascular disease models, we will summarise several of the above covariates by computing and adjusting for a cardiovascular risk score following the approach set out by van Staa et al (PLoS One. 2014 Oct 1;9(10):e106455.). We intend to use the QRISK2 approach, but if costs for the necessary commercial software make it impractical we will use the Framingham model.

We will also account for first vs subsequent MI, and for surgical intervention (percutaneous vs CABG vs none) in models and in the sensitivity analyses.

Direct measures of comorbidity or frailty will also be potentially included, from our recent work on diagnostic trends in the oldest old (Accepted Age and Ageing 2014), which modestly extends a previous multi-morbidity measure by Salisbury et al to be a little more representative of later life co-morbidity. These are based on a count of common diseases and conditions as well as "geriatric syndromes" such as falls and fractures, dizziness, incontinence and skin ulcers. The "Estimating Cardiovascular Risk in the Elderly" study will

also aim to elicit information about family history of cardiovascular disease, although limited recording of this is acknowledged.

**Patient or user group involvement**

This project is part of the NIHR School for Public health Research programme on 'Ageing Well'. In this framework, advancement of the work, partial and final results will be regularly shared and discussed in the context of advisory boards/meetings involving general practitioners, leaders from the academics, policy makers, and representatives from relevant patient and aging-related lobbying associations.

We anticipate widespread interest in the results, as there have been many papers and editorials in the journals pointing out the difficulties of prevention and treatment in various older groups in the absence of evidence of effectiveness.

We plan to publish in the public health, geriatrics and primary care journals and present at related conferences.

**Limitations of the study design, data sources, and analytic methods**

A fundamental limitation is that the original source of data is a database that was not originally created for research purposes. This is important, as characteristics that might allow post-hoc adjustment for non-randomised data are likely to have been unmeasured in routinely collected data. Here effects that are confounded with the intervention under study and the outcome are likely to give rise to biased estimates. However the purpose of this study is to deploy methods that can adjust for confounding, regardless of whether it has been measured. This includes all types of confounding, such as indication bias where, say, a history of CV disease has prompted vaccination. The method will not mitigate against that bias which, sometimes misleadingly described as confounding bias, has been imparted by the analytical method itself e.g: non-collapsibility when estimating conditional effects in non-linear models.

A second limitation is in the absence of widespread serological testing, the available outcomes to identify pneumococcal and influenza infections may be insensitive to other pathological causes, although we believe antibiotic prescriptions and, in the case of

pneumococcal pneumonia, hospitalisations, both qualified by symptom descriptions, offer the next best alternative.

Another important limitation is that PERR is only applicable to recurrent outcomes (such as angina or myocardial infarction) and only in secondary prevention settings. A consequence of this stipulation is that patients must be alive until the study period and death may impart further bias if used an outcome in the study period, or may be informative if used to censor survival times. Other quasi-experimental study designs, including instrumental variable estimation or marginal structural models, that are useful in primary prevention are sometimes difficult to be implemented in certain circumstances, such as in the analysis of survival times. In spite of the Yu's evidence of robustness to time and confounder interaction, an intrinsic assumption of the method is that the confounding effect is stable across the periods, for which it is adjusted. It should also be acknowledged from Yu's paper that current insight into the method suggests.

**Plans for disseminating and communicating study results**

We plan to publish the results of the effectiveness analyses for the influenza and pneumococcal vaccines against their primary respiratory outcomes as well as the issue of their effectiveness against major cardiovascular outcomes. In addition, we plan to publish academic articles on the development and extension of the quasi-experimental analytical methods used to address confounding bias. We believe the results will be informative for shaping policy on vaccination in older patients, and that the methodological work will help validate and make available new tools to assist researchers in strengthening causal inferences from electronic medical record data.

Other information:

The University of Exeter based part of the NIHR National School for Public Health Research programme is led by Prof Melzer, who has extensive experience of analysing observational data on older populations, including large databases. Professor William Henley will oversee the statistics: he currently leads an MRC methodology project grant on the Prior Event Rate

Ratio approach. Adam Streeter is an experienced statistician, who will be leading the analyses of the vaccination data, and will also be writing this up for a PhD.

Dr Alessandro Ble is a qualified geriatrician with extensive (>40 papers) experience in epidemiology of ageing. He has a Masters degree in medicines evaluation. Dr Jane Masoli is an academic clinical fellow in Geriatrics and will be leading work on the anti-hypertensives. Kirsty Bowman (MPH, PhD student) will be undertaking the modelling of risk factors in older people. Dr Mujica-Mota is a health economist with an interest in older populations, and has experience of analysing the GP Patient Survey with over 1m records.

Primary Care senior input will be from Professor Valderas (who has extensive related research experience especially on co-morbidity, plus clinical sessions in general practice) and Senior lecturer Dr Sue Richards.

Our existing CPRD analysis over the last two years has provided coding approaches to the risk factors, common diseases and geriatric syndromes of interest, with a paper on diagnostic trends in the oldest old accepted by the journal “Age and Ageing”. We have also gained experience on the coding of common prescriptions in older people. We have an 8 Terabyte server capable of supporting our planned analyses and extensive experience of large dataset analysis.

## ISAC protocol - Appendix A

### Pneumococcal vaccination coding in Clinical file

*medcode*: 11363, 30411, 36826

*immstype*: 13, 18, 28

*status*: 1

### Pneumococcal vaccination coding in Therapy file

*prodcode*: 821832, 42612, 42991

### Influenza vaccination coding in Clinical file

*medcode*: 6 9039 10821 12104 12336 18330 18684 21123 32942

35655 44555 94301 95092 97941 98047 98183 98184 98217 98234

98302 98303 98306 98449

*immstype*: 4

*status*: 1

### Influenza vaccination coding in Therapy file

*prodcode*: 398 639 834 922 1329 2139 2552 2601 9710

10030 11824 13595 16585 18612 27407 30156 30198 32391 38421

40760 40876

### Coding of selected commonly used antibiotics for respiratory infections ( list to be extended following British Thoracic society guidelines)

### Amoxicillin product codes from Therapy file:

9 48 62 133 427 503 585 847 870

1637 1722 1812 2153 2281 3669 3742 4154 7737 9243

11613 11634 12378 14371 14386 14396 14407 15148 17711 18786

21799 21827 21829 21844 21845 21963 22015 22016 22017 22415

22438 23238 23740 23967 24150 24200 24203 25484 26157 26262

27714 27725 28870 28872 28875 28882 29337 29463 29697 29858

30498 30528 30743 30745 31014 31286 31423 31535 31661 31801

32622 32640 32872 33109 33110 33112 33165 33222 33343 33570

33689 33690 33692 33696 33699 33706 34001 34042 34232 34384

34435 34638 34679 34714 34760 34775 34852 34855 34857 34885

34912 35570 36054 37755 38684 40238 40243 41090 41818 41835

**Doxycycline product codes from Therapy file:**

264 268 970 1046 2202 2884 3152 6396 8724 9267

10454 12987 14904 15071 21038 21828 21860 21878 23405 23432

23819 24126 24149 26392 26747 30739 32066 32419 33671 34175

34300 34423 34594 34765 40391 41560 41605 46807

**ICD-10 codes describing pneumonia in HES data:**

| **Description** | **Code** |
|---|---|
| Sepsis due to Streptococcus pneumoniae | A40.3 |
| Streptococcus pneumoniae as the cause of diseases classified to other chapters | B95.3 |
| Pneumococcal meningitis | G00.1 |
| Pneumonia due to Streptococcus pneumoniae | J13 |
| Pneumonia, organism unspecified | J18 |

**Coding of covariates**: we plan to use the same coding approaches as in our approved project 14_135R entitled " Estimating Cardiovascular risk in the Elderly." Below are the Read codes for coronary heart disease, for assessing the presence of disease in GP records. This will be used for patient selection but outcomes will be based on hospital HES records only and will include major CVD events.

| **Coronary heart disease (MI)** | | |
|---|---|---|
| MEDCODE | READCODE | CONDITION |
| 240 | G3...00 | Ischaemic heart disease |
| 241 | G30..00 | Acute myocardial infarction |
| 1204 | G30..14 | Heart attack |
| 1344 | G340.12 | Coronary artery disease |
| 1655 | G340.11 | Triple vessel disease of the heart |
| 1676 | G3z..00 | Ischaemic heart disease NOS |
| 1677 | G30..15 | MI - acute myocardial infarction |

| | | |
|---|---|---|
| 1678 | G308.00 | Inferior myocardial infarction NOS |
| 1792 | G3...13 | IHD - Ischaemic heart disease |
| 2491 | G30..12 | Coronary thrombosis |
| 3704 | G307.00 | Acute subendocardial infarction |
| 3999 | G340000 | Single coronary vessel disease |
| 4017 | G32..00 | Old myocardial infarction |
| 5254 | G340100 | Double coronary vessel disease |
| 5387 | G301.00 | Other specified anterior myocardial infarction |
| 5413 | G340.00 | Coronary atherosclerosis |
| 7320 | G343.00 | Ischaemic cardiomyopathy |
| 8935 | G302.00 | Acute inferolateral infarction |
| 9276 | G31y000 | Acute coronary insufficiency |
| 9413 | G31y.00 | Other acute and subacute ischaemic heart disease |
| 9507 | G307000 | Acute non-Q wave infarction |
| 9555 | G33z500 | Post infarct angina |
| 10562 | G307100 | Acute non-ST segment elevation myocardial infarction |
| 11983 | G311500 | Acute coronary syndrome |
| 12139 | G300.00 | Acute anterolateral infarction |
| 12229 | G30X000 | Acute ST segment elevation myocardial infarction |
| 13566 | G30..11 | Attack - heart |
| 13571 | G30..16 | Thrombosis - coronary |
| 14658 | G30z.00 | Acute myocardial infarction NOS |
| 14897 | G301z00 | Anterior myocardial infarction NOS |
| 14898 | G305.00 | Lateral myocardial infarction NOS |
| 15661 | G310.11 | Dressler's syndrome |

| | | |
|---|---|---|
| 15754 | G34z.00 | Other chronic ischaemic heart disease NOS |
| 16408 | G32..11 | Healed myocardial infarction |
| 17464 | G32..12 | Personal history of myocardial infarction |
| 17689 | G30..17 | Silent myocardial infarction |
| 17872 | G301100 | Acute anteroseptal infarction |
| 18842 | G35..00 | Subsequent myocardial infarction |
| 18889 | G34z000 | Asymptomatic coronary heart disease |
| 21844 | G31y300 | Transient myocardial ischaemia |
| 22383 | G3y..00 | Other specified ischaemic heart disease |
| 23078 | G34y100 | Chronic myocardial ischaemia |
| 23579 | G310.00 | Postmyocardial infarction syndrome |
| 23708 | G361.00 | Atrial septal defect/curr comp folow acut myocardal infarct |
| 23892 | G304.00 | Posterior myocardial infarction NOS |
| 24126 | G360.00 | Haemopericardium/current comp folow acut myocard infarct |
| 24540 | G34y000 | Chronic coronary insufficiency |
| 24783 | G3...11 | Arteriosclerotic heart disease |
| 25842 | G33z.00 | Angina pectoris NOS |
| 26863 | G33z600 | New onset angina |
| 27951 | G31..00 | Other acute and subacute ischaemic heart disease |
| 27977 | G31yz00 | Other acute and subacute ischaemic heart disease NOS |
| 28138 | G34..00 | Other chronic ischaemic heart disease |
| 28554 | G33zz00 | Angina pectoris NOS |
| 28736 | G30y000 | Acute atrial infarction |

| | | |
|---|---|---|
| 29421 | G344.00 | Silent myocardial ischaemia |
| 29553 | G366.00 | Thrombosis atrium,auric append&vent/curr comp foll acute MI |
| 29643 | G303.00 | Acute inferoposterior infarction |
| 29758 | G30X.00 | Acute transmural myocardial infarction of unspecif site |
| 29902 | G330z00 | Angina decubitus NOS |
| 30330 | G309.00 | Acute Q-wave infarct |
| 30421 | G30..13 | Cardiac rupture following myocardial infarction (MI) |
| 32272 | G38..00 | Postoperative myocardial infarction |
| 32450 | G33z400 | Ischaemic chest pain |
| 32854 | G30B.00 | Acute posterolateral myocardial infarction |
| 34328 | G311300 | Refractory angina |
| 34633 | G34y.00 | Other specified chronic ischaemic heart disease |
| 34803 | G30y.00 | Other acute myocardial infarction |
| 35713 | G34yz00 | Other specified chronic ischaemic heart disease NOS |
| 36423 | G36..00 | Certain current complication follow acute myocardial infarct |
| 36523 | G311.00 | Preinfarction syndrome |
| 36609 | G342.00 | Atherosclerotic cardiovascular disease |
| 37657 | G362.00 | Ventric septal defect/curr comp fol acut myocardal infarctn |
| 38609 | G351.00 | Subsequent myocardial infarction of inferior wall |
| 39449 | G312.00 | Coronary thrombosis not resulting in myocardial infarction |
| 39546 | Gyu3000 | [X]Other forms of angina pectoris |
| 39655 | G311.12 | Impending infarction |

| | | |
|---|---|---|
| 39693 | G31y200 | Subendocardial ischaemia |
| 40429 | G301000 | Acute anteroapical infarction |
| 41221 | G30y200 | Acute septal infarction |
| 41835 | G384.00 | Postoperative subendocardial myocardial infarction |
| 45809 | G350.00 | Subsequent myocardial infarction of anterior wall |
| 46017 | G30yz00 | Other acute myocardial infarction NOS |
| 46112 | G380.00 | Postoperative transmural myocardial infarction anterior wall |
| 46166 | G35X.00 | Subsequent myocardial infarction of unspecified site |
| 46276 | G381.00 | Postoperative transmural myocardial infarction inferior wall |
| 47637 | Gyu3300 | [X]Other forms of chronic ischaemic heart disease |
| 52517 | Gyu3.00 | [X]Ischaemic heart diseases |
| 54251 | G311z00 | Preinfarction syndrome NOS |
| 54535 | G33z100 | Stenocardia |
| 55137 | G311011 | MI - myocardial infarction aborted |
| 59189 | G363.00 | Ruptur cardiac wall w'out haemopericard/cur comp fol ac MI |
| 59940 | G364.00 | Ruptur chordae tendinae/curr comp fol acute myocard infarct |
| 61072 | G311000 | Myocardial infarction aborted |
| 62626 | G30y100 | Acute papillary muscle infarction |
| 63467 | G306.00 | True posterior myocardial infarction |
| 66388 | G33z000 | Status anginosus |
| 68357 | G31y100 | Microinfarction of heart |
| 68401 | Gyu3200 | [X]Other forms of acute ischaemic heart disease |
| 68748 | G38z.00 | Postoperative myocardial infarction, unspecified |

| | | |
|---|---|---|
| 69474 | G365.00 | Rupture papillary muscle/curr comp fol acute myocard infarct |
| 72562 | G353.00 | Subsequent myocardial infarction of other sites |
| 96838 | Gyu3400 | [X]Acute transmural myocardial infarction of unspecif site |
| 99991 | Gyu3600 | [X]Subsequent myocardial infarction of unspecified site |

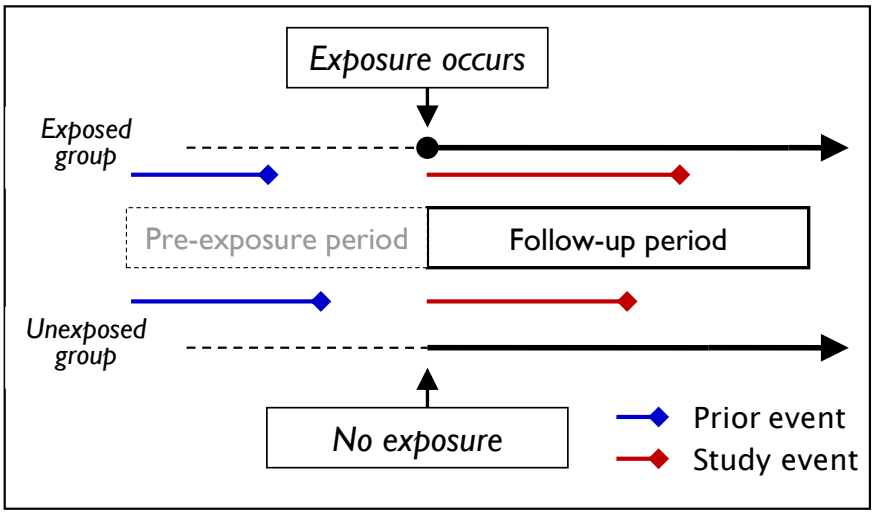


Figure 1: Timeline of prior and study observation periods for the PERR method

# Appendix D – CPRD and HES codes

**PPV codes (immstype) in immunisation file:**

13, 28

[both codes conditioned on status = 1, i.e: vaccine “given”]

**PPV codes (prodcode) in therapy file:**

821, 832, 42612, 42991

**Influenza vaccine codes in immunisation file:**

6, 9039, 10821, 12104, 12336, 18330, 18684, 21123, 32942, 35655, 44555, 94301, 95092, 97941, 98047, 98183, 98184, 98217, 98234, 98302, 98303, 98306, 98449

**Influenza vaccine codes in therapy file:**

398, 639, 834, 922, 1329, 2139, 2552, 2601, 9710, 10030, 11824, 13595, 16585, 18612, 27407, 30156, 30198, 32391, 38421, 40760, 40876

**ICD10 codes for hospitalisation for suspected pneumococcal pneumonia in HES data:**

J13, J15.8, J15.9, J16.8, J17, J18

**ICD10 codes for myocardial infarction admissions to hospital in HES data:**

I20.0, I21.0 - I21.4, I21.9, I22.0, I22.1, I22.8, I22.9, I21

**Amoxicillin codes in therapy file:**

9, 48, 62, 133, 427, 503, 585, 847, 870,  1637, 1722, 1812, 2153, 2281, 3669, 3742, 4154, 7737, 9243,  11613, 11634, 12378, 14371, 14386, 14396, 14407, 15148, 17711, 18786,  21799, 21827, 21829, 21844, 21845, 21963, 22015, 22016, 22017,

22415, 22438, 23238, 23740, 23967, 24150, 24200, 24203, 25484, 26157, 26262, 27714, 27725, 28870, 28872, 28875, 28882, 29337, 29463, 29697, 29858, 30498, 30528, 30743, 30745, 31014, 31286, 31423, 31535, 31661, 31801, 32622, 32640, 32872, 33109, 33110, 33112, 33165, 33222, 33343, 33570, 33689, 33690, 33692, 33696, 33699, 33706, 34001, 34042, 34232, 34384, 34435, 34638, 34679, 34714, 34760, 34775, 34852, 34855, 34857, 34885, 34912, 35570, 36054, 37755, 38684, 40238, 40243, 41090, 41818, 41835

**Doxycycline codes in therapy file:**

264, 268, 970, 1046, 2202, 2884, 3152, 6396, 8724, 9267, 10454, 12987, 14904, 15071, 21038, 21828, 21860, 21878, 23405, 23432, 23819, 24126, 24149, 26392, 26747, 30739, 32066, 32419, 33671, 34175, 34300, 34423, 34594, 34765, 40391, 41560, 41605, 46807

**Product codes in therapy file for antiviral drugs:**

| BNF code | CPRD prodcode | strength |
|---|---|---|
| Amantadine hydrochloride | 5339 | 100mg |
| Amantadine hydrochloride | 6035 | 50mg/5ml |
| Zanamivir | 6610 | 5mg |
| Oseltamivir phosphate | 10129 | 75mg |
| Oseltamivir phosphate | 10131 | 75mg |
| Oseltamivir phosphate | 10137 | 12mg/1ml |
| Oseltamivir phosphate | 18863 | 60mg/5ml |
| Zanamivir | 21169 | 5mg |
| Amantadine hydrochloride | 21745 | 50mg/5ml |
| Amantadine hydrochloride | 25890 | 100mg |
| Oseltamivir phosphate | 38523 | 30mg |
| Oseltamivir phosphate | 38955 | 30mg |
| Oseltamivir Phosphate | 39252 | 45mg |
| Oseltamivir phosphate | 39894 | 45mg |
| Oseltamivir phosphate | 40710 | 15mg/1ml |
| Oseltamivir phosphate | 42326 | 15mg/1ml |
| Oseltamivir phosphate | 52526 | 15mg/1ml |
| Oseltamivir phosphate | 53759 | 6mg/1ml |
| Oseltamivir phosphate | 54814 | 30mg/5ml |

**Medcodes from clinical file for symptom descriptions used to qualify antibiotic codes:**

| Medcode | Description |
|---|---|
| 293 | Respiratory tract infection |
| 4899 | Recurrent chest infection |
| 68 | Chest infection |
| 2581 | Chest infection NOS |
| 3358 | Lower resp tract infection |
| 5534 | Pneumococcal infection |
| 7074 | Respiratory infection NOS |
| 8025 | Acute respiratory infections |
| 14804 | Sputum appears infected |
| 16287 | Chest infection - unspecified bronchopneumonia |
| 17359 | Chest infection - unspecified bronchitis |
| 19400 | Chest infection - pnemonia due to unspecified organism |
| 21061 | Chronic obstruct pulmonary dis with acute lower resp infectn |
| 21113 | Acute respiratory infection NOS |
| 22795 | Chest infection - other bacterial pneumonia |
| 23640 | Other specified acute respiratory infections |
| 3382 | Streptococcal infection |
| 572 | Pneumonia due to unspecified organism |
| 886 | Bronchopneumonia due to unspecified organism |
| 1849 | Lobar (pneumococcal) pneumonia |
| 3683 | Basal pneumonia due to unspecified organism |

| | |
|---|---|
| 9639 | Lobar pneumonia due to unspecified organism |
| 10086 | Pneumonia and influenza |
| 11849 | Other specified pneumonia or influenza |
| 12423 | Pneumonia due to streptococcus |
| 13573 | Influenza with bronchopneumonia |
| 22009 | Streptococ pneumon/cause/disease classified/oth chapters |
| 23095 | Bacterial pneumonia NOS |
| 25694 | Pneumonia due to other specified organisms |
| 23333 | Hypostatic pneumonia |
| 24356 | Hypostatic bronchopneumonia |
| 1934 | Laryngotracheobronchitis |
| 1019 | Acute bronchiolitis |
| 17185 | Acute bronchiolitis with bronchospasm |
| 17917 | Acute bronchiolitis NOS |
| 29669 | Acute bronchitis and bronchiolitis |
| 41137 | Acute bronchitis or bronchiolitis NOS |
| 2195 | Bronchiectasis |
| 20364 | Recurrent bronchiectasis |
| 1234 | Productive cough NOS |
| 7708 | Productive cough-yellow sputum |
| 7773 | Productive cough -green sputum |
| 18907 | Cough with fever |
| 8760 | [D]Positive culture findings in sputum |
| 15430 | [D]Sputum abnormal - colour |

| | |
|---|---|
| 16026 | Sputum examination: abnormal |
| 24181 | Sputum: mucopurulent |
| 30754 | Yellow sputum |
| 36880 | Green sputum |

# Appendix E – Diagnostic plots for Cox models in influenza study

As a visual inspection of the extent to which the Cox models of MIs fitted to vaccination status, adjusting for age and gender, may have deviated from the proportional hazards assumption, the negative logarithm of the hazard function (also expressed as -ln(-ln(Survival probability)) in the plots) was plotted against the natural logarithm of analysis time in Stata – so-called log-log plots. The nature of transformation meant the most data points lay to the right end of the x-axis corresponding to the majority of survival times, which were greater than about 50 days (ln(50) ≈ 4). Apart from the sparse points in the plots corresponding to the shorter survival times, the bulk of the data seemed to produce reasonably parallel lines for the vaccine recipients and controls. There was some concern about the study period of the 2009 and 1998 cohorts, and the prior period of the 1997 cohorts. The lines of their log-log plots were the least parallel of all the plots, and thus, appeared to potentially indicate deviation from the proportional hazards assumption. However, the estimates for these particular periods did not appear incongruous or remarkable, and so the condition of proportional hazards was assumed to be broadly satisfied.

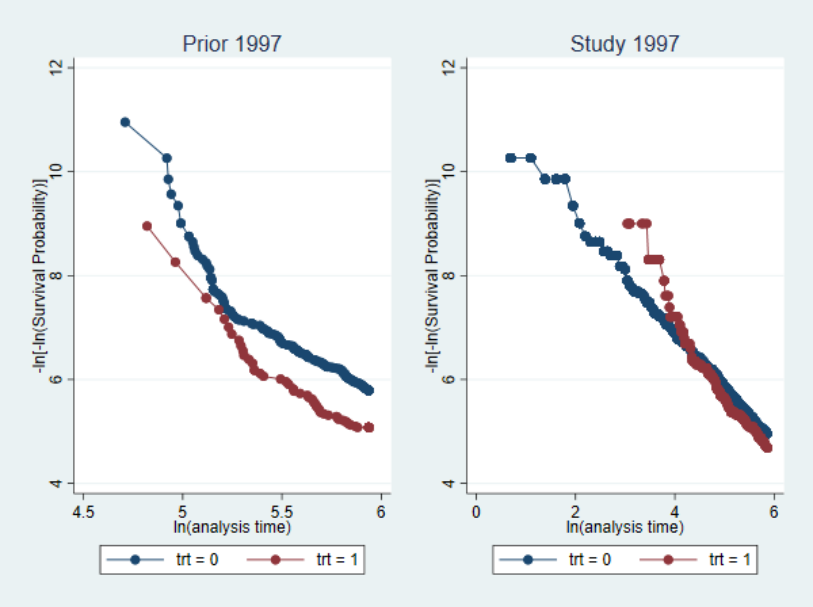


*Figure 42:* log-log plot for the prior and study period Cox model of MIs on vaccination status in the 1997 cohort adjusted for age and gender

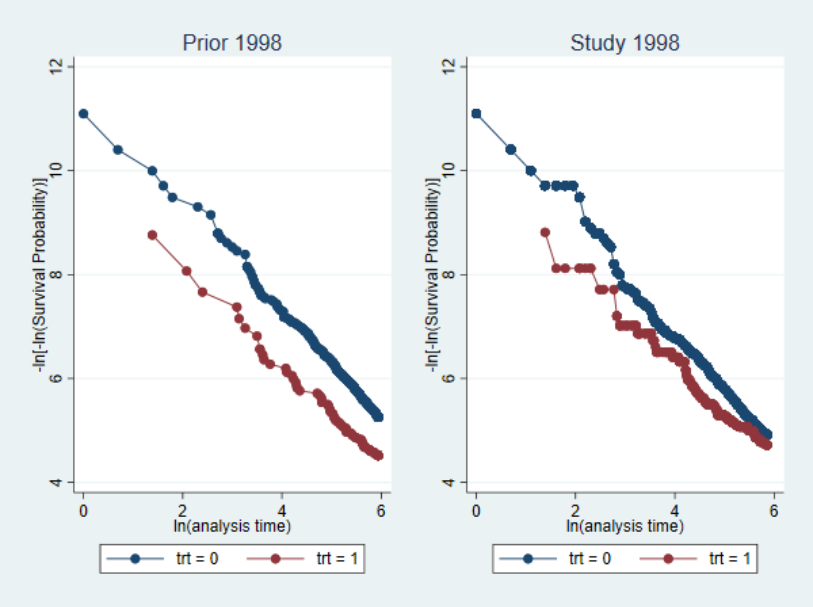


*Figure 43:* log-log plot for the prior and study period Cox model of MIs on vaccination status in the 1998 cohort adjusted for age and gender

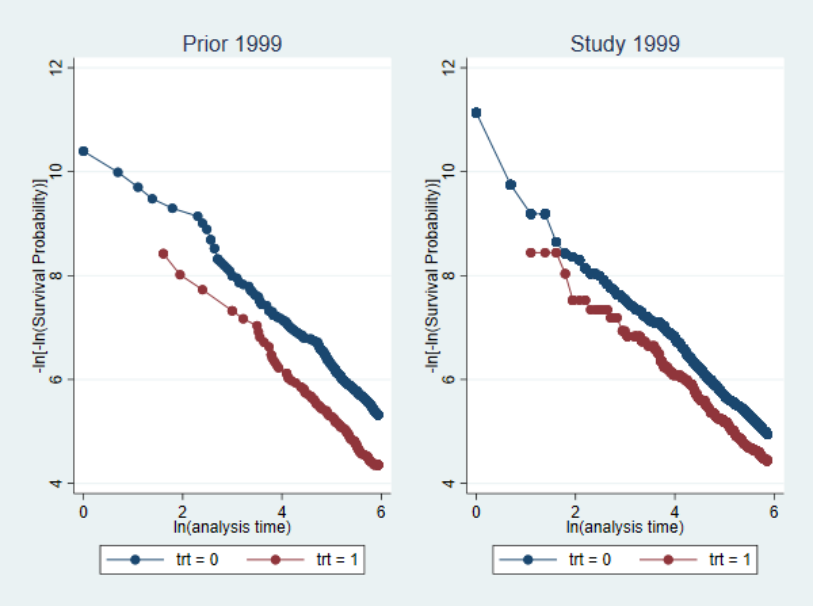


*Figure 44:* log-log plot for the prior and study period Cox model of MIs on vaccination status in the 1999 cohort adjusted for age and gender

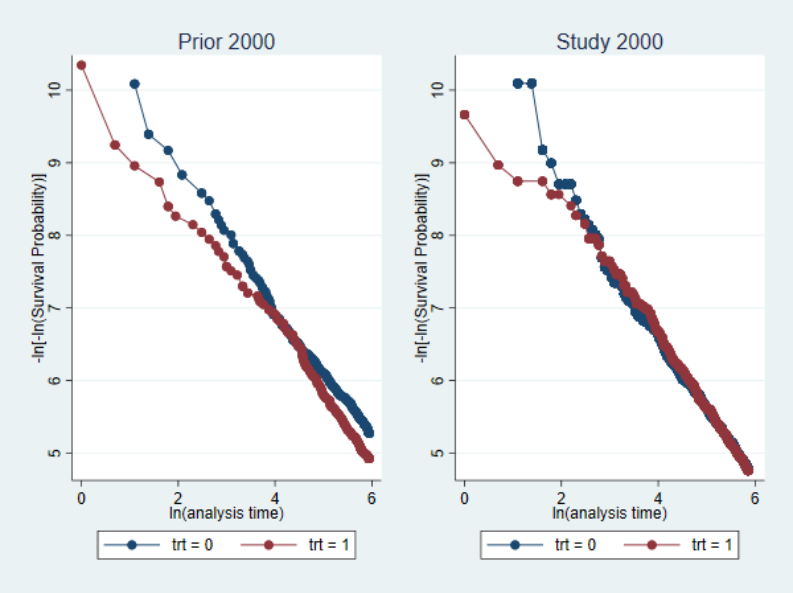


*Figure 45:* log-log plot for the prior and study period Cox model of MIs on vaccination status in the 2000 cohort adjusted for age and gender

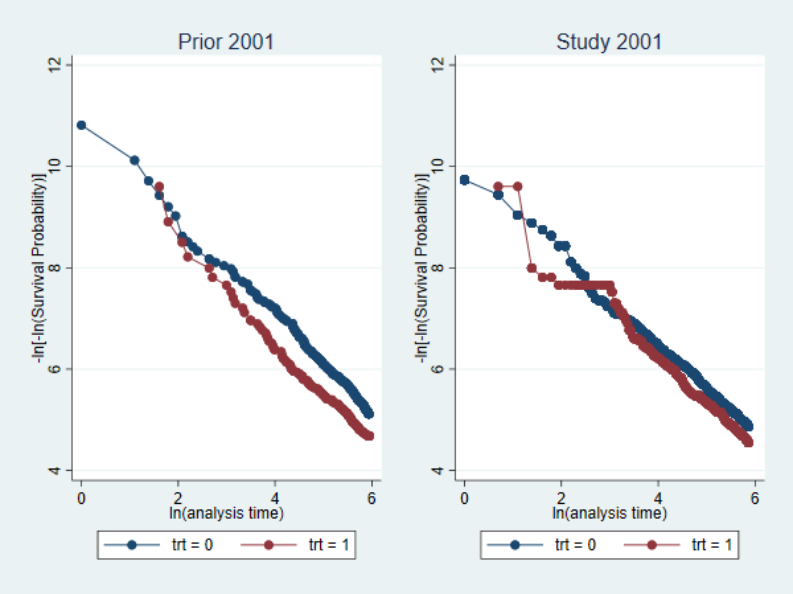


*Figure 46:* log-log plot for the prior and study period Cox model of MIs on vaccination status in the 2001 cohort adjusted for age and gender

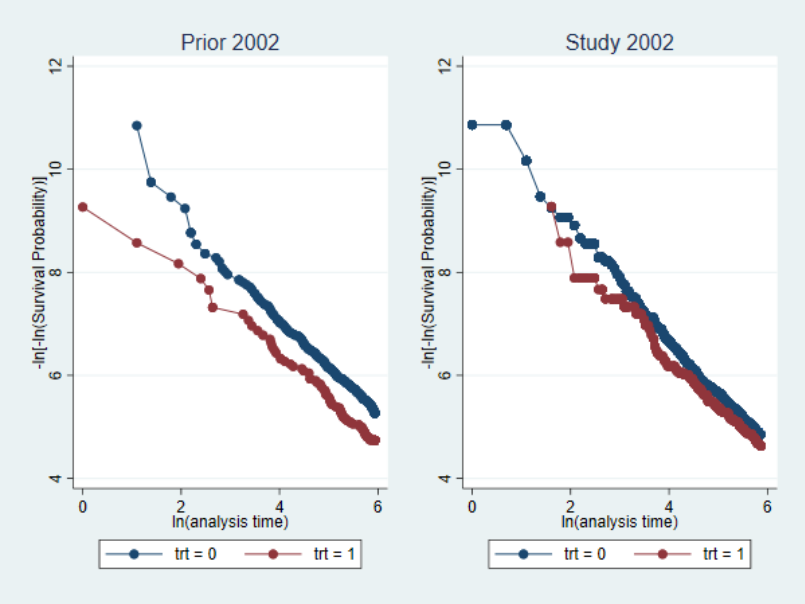


*Figure 47:* log-log plot for the prior and study period Cox model of MIs on vaccination status in the 2002 cohort adjusted for age and gender

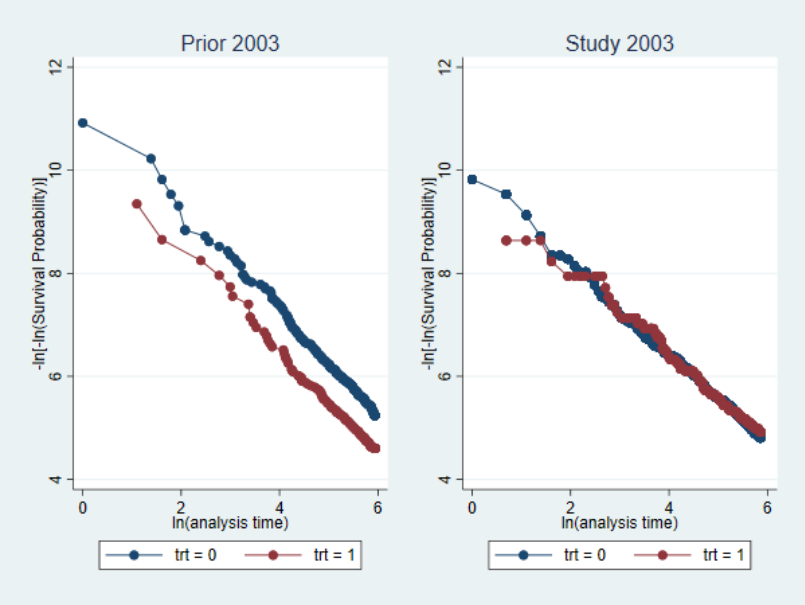


*Figure 48:* log-log plot for the prior and study period Cox model of MIs on vaccination status in the 2003 cohort adjusted for age and gender

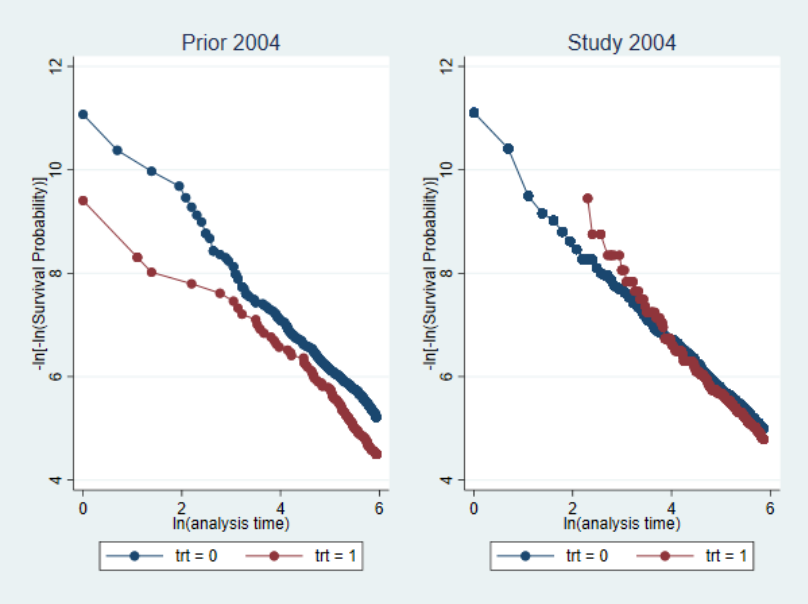


*Figure 49:* log-log plot for the prior and study period Cox model of MIs on vaccination status in the 2004 cohort adjusted for age and gender

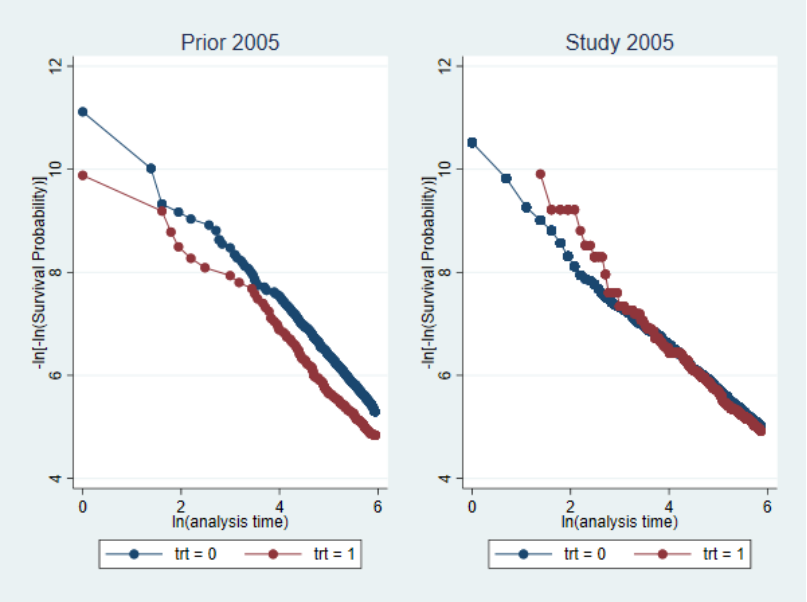


*Figure 50:* log-log plot for the prior and study period Cox model of MIs on vaccination status in the 2005 cohort adjusted for age and gender

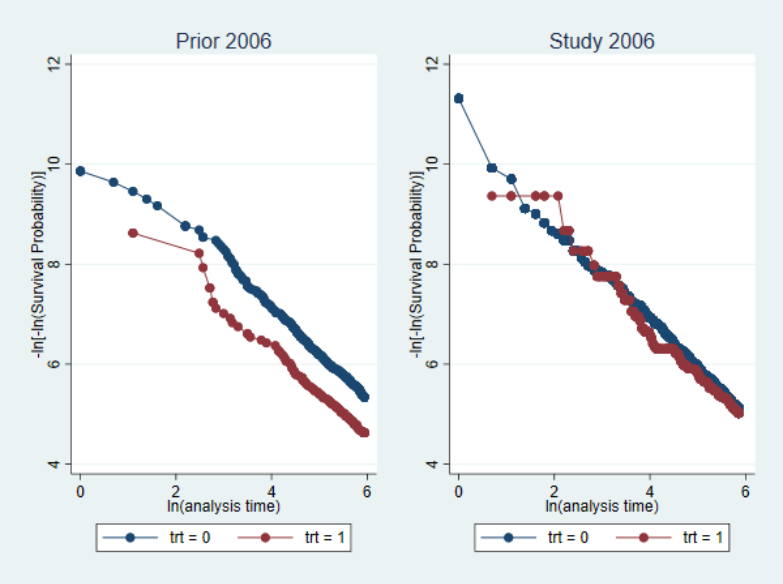


*Figure 51:* log-log plot for the prior and study period Cox model of MIs on vaccination status in the 2006 cohort adjusted for age and gender

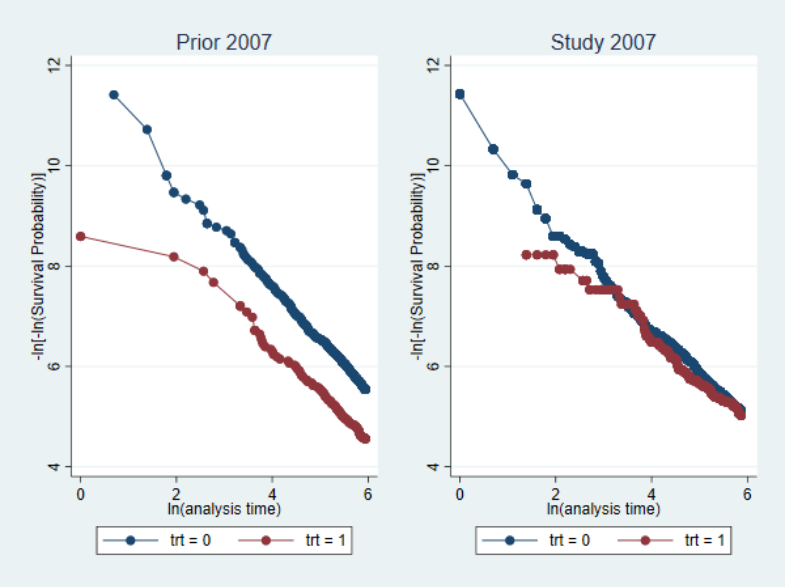


*Figure 52:* log-log plot for the prior and study period Cox model of MIs on vaccination status in the 2007 cohort adjusted for age and gender

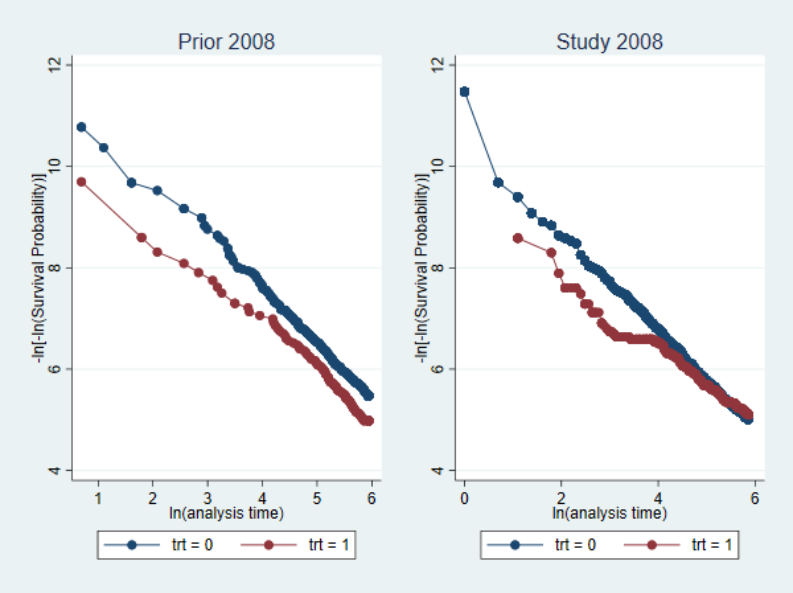


*Figure 53:* log-log plot for the prior and study period Cox model of MIs on vaccination status in the 2009 cohort adjusted for age and gender

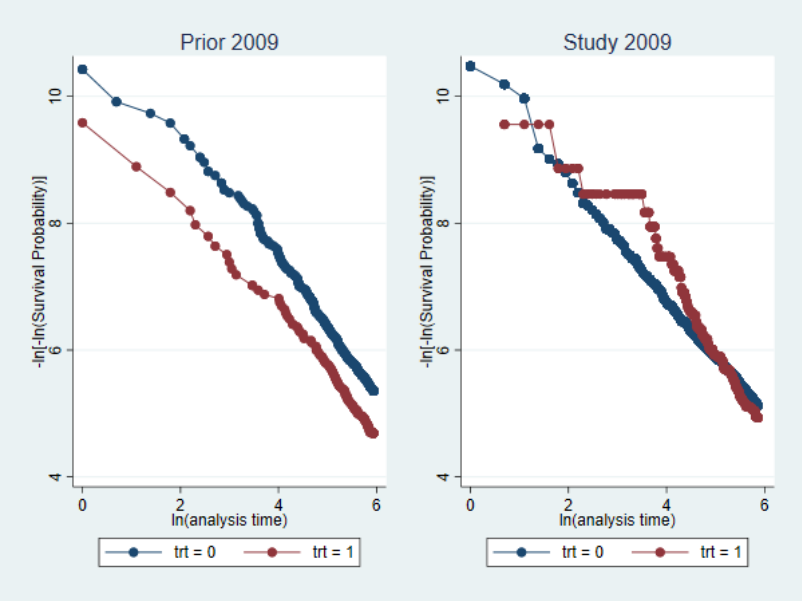


*Figure 54:* log-log plot for the prior and study period Cox model of MIs on vaccination status in the 2009 cohort adjusted for age and gender

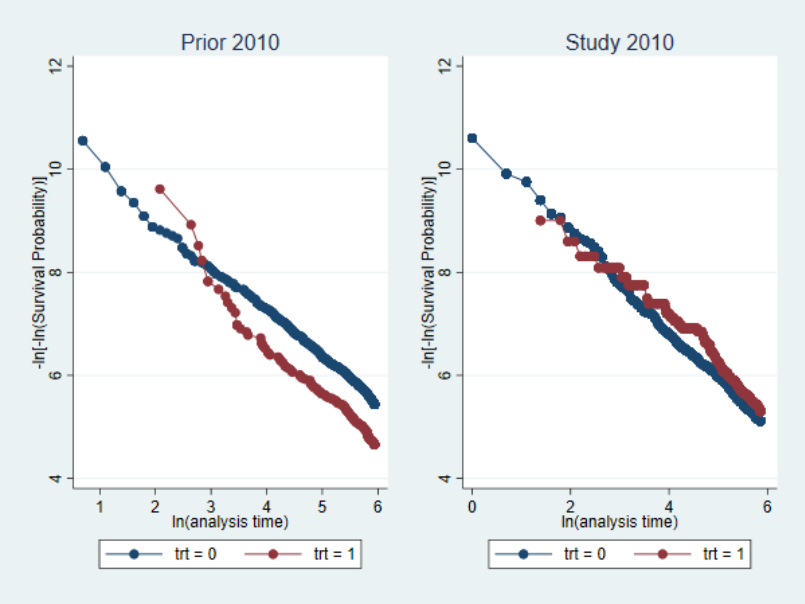


*Figure 55:* log-log plot for the prior and study period Cox model of MIs on vaccination status in the 2010 cohort adjusted for age and gender

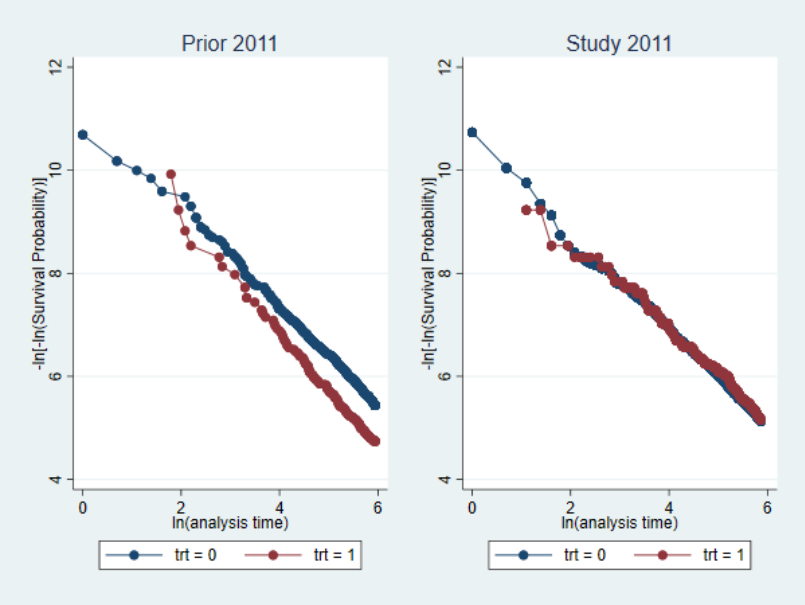


*Figure 56:* log-log plot for the prior and study period Cox model of MIs on vaccination status in the 2011 cohort adjusted for age and gender

# Appendix F – codes for statins in CPRD data

### Atorvastatin

28 75 745 2955 3411 5775 7374 17683 47065 47090
47630 47721 48518 49558 49751 5023 50272 50788 50790
50963 51134 51200 51359 51622 51876 52097 52168 52211
52397 52398 52459 52460 52821 53594 53772 53887 53890
54535 55032 55034 55444 55727 56182 56248 56564 56841
57117 57348 57834 57836 58041 58110 58394 58418 58742
58834 58868 59272 59331 59357 59446 59776 59859 60511
60607 60989 61149

### Rosuvastatin

713 6213 7347 7554 9897 9930 15252 17688 53460
57763 57999 59447 59452 60160

### Fluvastatin

379 2137 5985 8380 9153 11627 53770 59278

### Simvastatin

25 42 51 802 818 2718 5148 6168
9920 13041 22579 31930 32909 33082 34312 34316
34353 34366 34376 34381 34476 34481 34502 34535 34545
34560 34746 34814 34879 34891 34907 34955 34969 37434
39060 39652 39675 39870 40340 40601 41657 44528 44650
44878 45219 45235 45245 45346 46878 46956 47774 47948
48018 48051 48058 48078 48431 48867 49061 49062 49587
50483 50564 50670 50703 50754 50882 51085 51166 51233
51483 51715 52098 52257 52625 52676 52812 52953 52962
53087 53340 53415 53676 53822 53908 53966 54240 54266
54493 54655 54819 54947 54976 54985 55452 56481 56494
57568 58315 58755 61155 61321 61360 61665

**Pravastatin**

490 730 1219 1221 1223 3690

32921 34820 36377 40382 43218 47988 48097 50925 51676

51890 52755 54435 54607 55912 56146 56607 56735 56893

56916 57108 57137 57296 57397 59508 60251 61134

**Simvastatin + ezetimibe**

7552 10172 10183 10206 11815 14219 16186 17059 21020

# References


1. Lewis D. Causation. *J Philos*. 1973;70(17):556-567. doi:10.2307/2025310.

2. Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Health*. 2001;22:189-212. doi:10.1146/annurev.publhealth.22.1.189.

3. Splawa-Neyman J, Dabrowska DM, Speed TP. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Stat Sci*. 1990;5(4):465-472. doi:10.2307/2245382.

4. Rubin DB. Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Stat Sci*. 1990;5(4):472-480. doi:10.2307/2245383.

5. Greenland S. Causal Analysis in the Health Sciences. *J Am Stat Assoc*. 2000;95(449):286-289. http://www.jstor.org/stable/2669548. Accessed December 12, 2017.

6. Glass TA, Goodman SN, Hernán MA, Samet JM. Causal Inference in Public Health. *Annu Rev Public Health*. 2013;34(1):61-75. doi:10.1146/annurev-publhealth-031811-124606.

7. Dawid AP. Causal inference without counterfactuals. *J Am Stat Assoc*. 2000;95(450):407-424. doi:10.1080/01621459.2000.10474210.

8. Rubin DB. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *J Educ Psychol*. 1974;66(5):688-701. http://eric.ed.gov/?id=EJ118470. Accessed December 27, 2013.

9. Holland PW. Statistics and Causal Inference. *J Am Stat Assoc*. 1986;81(396):945-960. doi:10.1080/01621459.1986.10478354.

10. Cochran W, Rubin D. Controlling bias in observational studies: A review. *Sankhyā Indian J Stat Ser A*. 1973;35(4):417-446. http://www.jstor.org/stable/10.2307/25049893. Accessed February 6, 2014.

11. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55. doi:10.1093/biomet/70.1.41.

12. Robins J, Hernán M, Brumback B. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology*. 2000;11(5):550-560. http://www.jstor.org/stable/3703997. Accessed November 8, 2015.

13. Morgan SL, Todd JJ. A diagnostic routine for the detection of consequential heterogeneity of causal effects. *Sociol Methodol*. 2008;38(1):231-281. doi:10.1111/j.1467-9531.2008.00204.x.

14. Hernán MA. A definition of causal effect for epidemiological research. *J Epidemiol Community Health*. 2004;58(4):265-271. doi:10.1136/jech.2002.006361.

15. Greenland S, Robins JM. Identifiability, Exchangeability, and Epidemiological Confounding. *Int J Epidemiol*. 1986;15(3):413-419. doi:10.1093/ije/15.3.413.

16. Hernán MA. Does water kill? A call for less casual causal inferences. *Ann Epidemiol*. 2016;26(10):674-680. doi:10.1016/j.annepidem.2016.08.016.

17. Cox DR. *Planning of Experiments. Wiley*. New York, NY: Wiley; 1958.

18. Rubin DB. Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment. *Source J Am Stat Assoc*. 1980;75(371):591-593. http://www.jstor.org/stable/2287653. Accessed January 14, 2018.

19. Imai K, Tingley D, Yamamoto T. Experimental designs for identifying causal mechanisms. *J R Stat Soc Ser A (Statistics Soc*. 2013;176(1):5-51. doi:10.1111/j.1467-985X.2012.01032.x.

20. Simpson EH. The Interpretation of Interaction in Contingency Tables. *J R Stat Soc Ser B*. 1951;13(2):238-241. doi:10.1038/203024b0.

21. Brookhart MA, Stürmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. *Med Care*. 2010;48(6 Suppl):S114-20. doi:10.1097/MLR.0b013e3181dbebe3.

22. Brooks JM, Ohsfeldt RL. Squeezing the balloon: propensity scores and unmeasured covariate balance. *Health Serv Res*. 2013;48(4):1487-1507. doi:10.1111/1475-6773.12020.

23. Wright PG. *The Tariff on Animal and Vegetable Oils*. New York, NY: The

Macmillan Co.; 1928. https://scholar.google.co.uk/scholar?as_q=Tariff+on+animal+and+vegetable+oils&as_epq=&as_oq=&as_eq=&as_occt=any&as_sauthors=wright&as_publication=&as_ylo=&as_yhi=&btnG=&hl=en&as_sdt=0%2C5#6. Accessed January 12, 2015.

24. Ashenfelter O, Card D. Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs. *Rev Econ Stat*. 1985;67(4):648-660. http://www.jstor.org/stable/1924810. Accessed May 26, 2016.

25. Thistlethwaite, Donald L. Campbell DT. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *J Educ Psychol*. 1960;51(6):309-317. doi:10.1037/h0044319.

26. Cook TD, Campbell DT. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. 3rd ed. Chicago: Rand McNally, 1979; 1979. http://lib.exeter.ac.uk/record=b1090074~S6. Accessed January 13, 2015.

27. Pearl J. *Causality : Models, Reasoning, and Inference*. Cambridge University Press; 2000. https://books.google.co.uk/books/about/Causality.html?id=wnGU_TsW3BQC&redir_esc=y. Accessed December 20, 2017.

28. Girshick, M. A.; Trygve H. Statistical Analysis of the Demand for Food: Examples of Simultaneous Estimation of Structural Equations. *Econometrica*. 1947;15(2):79-110. doi:10.2307/1907066.

29. Pearl J. Causal inference in statistics: An overview *. *Stat Surv*. 2009;3:96-146. doi:10.1214/09-SS057.

30. Aalen OO, Roysland K, Gran JM, Kouyos R, Lange T. Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms. *Stat Methods Med Res*. 2014;25(5):2294-2314. doi:10.1177/0962280213520436.

31. Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: The need for a pluralistic approach. *Int J Epidemiol*. 2016;45(6):1776-1786. doi:10.1093/ije/dyv341.

32. Schwartz S, Gatto NM, Campbell UB. Causal identification: a charge of epidemiology in danger of marginalization. *Ann Epidemiol*. 2016;26(10):669-673. doi:10.1016/j.annepidem.2016.03.013.

33. Maldonado G. The role of counterfactual theory in causal reasoning. *Ann Epidemiol*. 2016;26(10):681-682. doi:10.1016/j.annepidem.2016.08.017.

34. Daniel RM, De Stavola BL, Vansteelandt S. Commentary: The formal approach to quantitative causal inference in epidemiology: Misguided or misrepresented? *Int J Epidemiol*. 2016;45(6):1817-1829. doi:10.1093/ije/dyw227.

35. Krieger N, Smith GD. The tale wagged by the DAG: Broadening the scope of causal inference and explanation for epidemiology. *Int J Epidemiol*. 2016;45(6):1787-1808. doi:10.1093/ije/dyw114.

36. Aalen OO, Røysland K, Gran JM, Ledergerber B. Causality, mediation and time: a dynamic viewpoint. *J R Stat Soc Ser A Stat Soc*. 2012;175(4):831-861. doi:10.1111/j.1467-985X.2011.01030.x.

37. Kim J, Jung H, Bates DW. History and trends of “personal health record” research in PubMed. *Healthc Inform Res*. 2011;17(1):3-17. doi:10.4258/hir.2011.17.1.3.

38. Clinical Practice Research Datalink - CPRD. https://www.cprd.com/home/. Accessed December 23, 2017.

39. Lawson DH, Sherman V, Hollowell J. The General Practice Research Database. *QJM An Int J Med*. 1998;91(6):445-452. doi:https://doi.org/10.1093/qjmed/91.6.445.

40. Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. *Ther Adv Drug Saf*. 2012;3(2):89-99. doi:10.1177/2042098611435911.

41. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015;44(3):827-836. doi:10.1093/ije/dyv098.

42. Woodcock A, Bakerly ND, New JP, et al. The Salford Lung Study protocol: A pragmatic, randomised phase III real-world effectiveness trial in asthma. *BMC Pulm Med*. 2015;15(160):1-5. doi:10.1186/s12890-015-0150-8.

43. Woodcock A, Vestbo J, Bakerly ND, et al. Effectiveness of fluticasone furoate plus vilanterol on asthma control in clinical practice: an open-label, parallel group, randomised controlled trial. *Lancet*. 2017;390(10109):2247-2255. doi:10.1016/S0140-6736(17)32397-8.

44. Dregan A, Van Staa TP, McDermott L, et al. Point-of-care cluster randomized trial in stroke secondary prevention using electronic health records. *Stroke*. 2014;45(7):2066-2071. doi:10.1161/STROKEAHA.114.005713.

45. Gulliford MC, van Staa T, Dregan A, et al. Electronic health records for intervention research: a cluster randomized trial to reduce antibiotic prescribing in primary care (eCRT study). *Ann Fam Med*. 2014;12(4):344-351. doi:10.1370/afm.1659.

46. van Staa T-P, Dyson L, McCann G, et al. The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials. *Health Technol Assess (Rockv)*. 2014;18(43):1-146. doi:10.3310/hta18430.

47. Staa T-P van, Goldacre B, Gulliford M, et al. Pragmatic randomised trials using routine electronic health records: putting them to the test. *BMJ*. 2012;344(e55):1-7. doi:10.1136/bmj.e55.

48. Gulliford MC, van Staa TP, McDermott L, McCann G, Charlton J, Dregan A. Cluster randomized trials utilizing primary care electronic health records: methodological issues in design, conduct, and analysis (eCRT Study). *Trials*. 2014;15(1):220. doi:10.1186/1745-6215-15-220.

49. Bunn S, Crane J. POSTnotes: Electronic Health Records. Number 519. 2016. http://researchbriefings.parliament.uk/ResearchBriefing/Summary/POST-PN-0519#fullreport. Accessed October 15, 2017.

50. *Five Year Forward View*.; 2014. https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf. Accessed October 15, 2017.

51. Corbett J, D'Angelo C, Gangitano L, Freeman J. *Future of Health: Findings from a Survey of Stakeholders on the Future of Health and Healthcare in England.* Santa Monica, CA: RAND Corporation; 2017. doi:10.7249/RR2147.

52. Shephard E, Stapley S, Hamilton W. The use of electronic databases in primary care research. *Fam Pract*. 2011;28(4):352-354. doi:10.1093/fampra/cmr039.

53. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015;44(3):827-836. doi:10.1093/ije/dyv098.

54. Collier R. Legumes, lemons and streptomycin: a short history of the clinical trial. *CMAJ*. 2009;180(1):23-24. doi:10.1503/cmaj.081879.

55. Bhatt A. Evolution of clinical research: a history before and beyond james lind. *Perspect Clin Res*. 2010;1(1):6-10. http://www.ncbi.nlm.nih.gov/pubmed/21829774. Accessed December 30, 2017.

56. Brown B, Smeeth L, van Staa T, Buchan I. Better care through better use of data in GP-patient partnerships. *Br J Gen Pract*. 2017;67(655):54-55. doi:10.3399/bjgp17X688921.

57. Gulliford MC, Moore M V, Little P, et al. Safety of reduced antibiotic prescribing for self limiting respiratory tract infections in primary care: cohort study using electronic health records. *BMJ*. 2016;354354:1-10. doi:10.1136/bmj.i3410.

58. Bartlett C, Doyal L, Ebrahim S, et al. The causes and effects of socio-demographic exclusions from clinical trials. *Health Technol Assess*. 2005;9(38):iii-iv, ix-x, 1-152. http://www.ncbi.nlm.nih.gov/pubmed/16181564. Accessed October 6, 2014.

59. Golomb BA, Chan VT, Evans MA, Koperski S, White HL, Criqui MH. The older the better: are elderly study participants more non-representative? A cross-sectional analysis of clinical trial and observational study samples. *BMJ Open*. 2012;2(6):e000833-. doi:10.1136/bmjopen-2012-000833.

60. Melzer D, Tavakoly B, Winder RE, et al. Much more medicine for the oldest

old: trends in UK electronic clinical records. *Age Ageing*. 2015;44(1):46-53. doi:10.1093/ageing/afu113.

61. Ble A, Masoli JAH, Barry HE, et al. Any versus long-term prescribing of high risk medications in older people using 2012 Beers Criteria: results from three cross-sectional samples of primary care records for 2003/4, 2007/8 and 2011/12. *BMC Geriatr*. 2015;15(1):146. doi:10.1186/s12877-015-0143-8.

62. Ble A, Hughes PM, Delgado J, et al. Safety and Effectiveness of Statins for Prevention of Recurrent Myocardial Infarction in 12 156 Typical Older Patients: A Quasi-Experimental Study. *Journals Gerontol Ser A Biol Sci Med Sci*. 2017;72(2):243-250. doi:10.1093/gerona/glw082.

63. Jackson LA, Nelson JC, Benson P, et al. Functional status is a confounder of the association of influenza vaccine and risk of all cause mortality in seniors. *Int J Epidemiol*. 2006;35(2):345-352. doi:10.1093/ije/dyi275.

64. Fedarko NS. The Biology of Aging and Frailty. *Clin Geriatr Med*. 2011;27(1):27-37. doi:10.1016/j.cger.2010.08.006.

65. Heuberger RA. The Frailty Syndrome: A Comprehensive Review. *J Nutr Gerontol Geriatr*. 2011;30(4):315-368. doi:10.1080/21551197.2011.623931.

66. Xue QL. The Frailty Syndrome: Definition and Natural History. *Clin Geriatr Med*. 2011;27(1):1-15. doi:10.1016/j.cger.2010.08.009.

67. Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. *Lancet*. 2013;381(9868):752-762. doi:10.1016/S0140-6736(12)62167-9.

68. Buchman AS, Wilson RS, Bienias JL, Bennett DA. Change in frailty and risk of death in older persons. *Exp Aging Res*. 2009;35(1):61-82. doi:10.1080/03610730802545051.

69. Fulop T, Larbi A, Witkowski JM, et al. Aging, frailty and age-related diseases. *Biogerontology*. 2010;11(5):547-563. doi:10.1007/s10522-010-9287-2.

70. McElhaney JE. Influenza vaccine responses in older adults. *Ageing Res Rev*. 2011;10(3):379-388. doi:10.1016/j.arr.2010.10.008.

71. Goronzy JJ, Weyand CM. Understanding immunosenescence to improve

responses to vaccines. *Nat Immunol*. 2013;14(5):428-436. doi:10.1038/ni.2588.

72. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA*. 2013;309(13):1351-1352. doi:10.1001/jama.2013.393.

73. Safran C, Bloomrosen M, Hammond WE, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *J Am Med Inform Assoc*. 2007;14(1):1-9. doi:10.1197/jamia.M2273.

74. Lund JL, Richardson DB, Stürmer T. The Active Comparator, New User Study Design in Pharmacoepidemiology: Historical Foundations and Contemporary Application. *Curr Epidemiol Reports*. 2015;2(4):221-228. doi:10.1007/s40471-015-0053-5.

75. Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. *J Clin Epidemiol*. 2005;58(6):550-559. doi:10.1016/j.jclinepi.2004.10.016.

76. Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol*. 2006;59(5):437-447. doi:10.1016/j.jclinepi.2005.07.004.

77. Stürmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am J Epidemiol*. 2005;162(3):279-289. doi:10.1093/aje/kwi192.

78. Stürmer T, Schneeweiss S, Rothman KJ, Avorn J, Glynn RJ. Performance of propensity score calibration - A simulation study. *Am J Epidemiol*. 2007;165(10):1110-1118. doi:http://dx.doi.org/10.1093/aje/kwm074.

79. Lunt M, Glynn RJ, Rothman KJ, Avorn J, Stürmer T. Propensity score calibration in the absence of surrogacy. *Am J Epidemiol*. 2012;175(12):1294-1302. doi:10.1093/aje/kwr463.

80. Lee W-C. Detecting and correcting the bias of unmeasured factors using perturbation analysis: a data-mining approach. *BMC Med Res Methodol*. 2014;14(1):18. doi:10.1186/1471-2288-14-18.

81. Brunner EJ, Kivimaki M, Witte DR, et al. Inflammation, insulin resistance, and diabetes--Mendelian randomization using CRP haplotypes points upstream. *PLoS Med*. 2008;5(8):e155. internal-pdf://137.108.247.40/Infl_diabet_MendRand-Brunner-PLoS08.pdf.

82. Burgess S, Thompson SG. Avoiding bias from weak instruments in mendelian randomization studies. *Int J Epidemiol*. 2011;40(3):755-764. doi:http://dx.doi.org/10.1093/ije/dyr036.

83. Wehby GL, Scholder S. Genetic instrumental variable studies of effects of prenatal risk factors. *Biodemography Soc Biol*. 2013;59(1):4-36. doi:http://dx.doi.org/10.1080/19485565.2013.774615.

84. Haring R, Teumer A, Völker U, et al. Mendelian randomization suggests non-causal associations of testosterone with cardiometabolic risk factors and mortality. *Andrology*. 2013;1(1):17-23. doi:10.1111/j.2047-2927.2012.00002.x.

85. Jokela M, Elovainio M, Keltikangas-Jarvinen L, et al. Body mass index and depressive symptoms: Instrumental-variables regression with genetic risk score. *Genes, Brain Behav*. 2012;11(8):942-948. doi:http://dx.doi.org/10.1111/j.1601-183X.2012.00846.x.

86. Kivimaki M, Magnussen CG, Juonala M, et al. Conventional and Mendelian randomization analyses suggest no association between lipoprotein(a) and early atherosclerosis: The Young Finns Study. *Int J Epidemiol*. 2011;40(2):470-478. doi:http://dx.doi.org/10.1093/ije/dyq205.

87. Laschkolnig A, Kollerits B, Lamina C, et al. Lipoprotein (a) concentrations, apolipoprotein (a) phenotypes, and peripheral arterial disease in three independent cohorts. *Cardiovasc Res*. 2014;103(1):28-36. doi:http://dx.doi.org/10.1093/cvr/cvu107.

88. Lawlor DA, Harbord RM, Timpson NJ, et al. The association of C-reactive protein and CRP genotype with coronary heart disease: findings from five

studies with 4,610 cases amongst 18,637 participants. *PLoS ONE [Electronic Resour*. 2008;3(8):e3011. doi:http://dx.doi.org/10.1371/journal.pone.0003011.

89. Leong A, Rehman W, Dastani Z, et al. The Causal Effect of Vitamin D Binding Protein (DBP) Levels on Calcemic and Cardiometabolic Diseases: A Mendelian Randomization Study. *PLoS Med*. 2014;11(10):e1001751. doi:http://dx.doi.org/10.1371/journal.pmed.1001751.

90. Nimptsch K, Aleksandrova K, Boeing H, et al. Association of CRP genetic variants with blood concentrations of C-reactive protein and colorectal cancer risk. *Int J Cancer*. 2015;136(5):1181-1192. doi:http://dx.doi.org/10.1002/ijc.29086.

91. Palmer TM, Sterne JAC, Harbord RM, et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in mendelian randomization analyses. *Am J Epidemiol*. 2011;173(12):1392-1403. doi:http://dx.doi.org/10.1093/aje/kwr026.

92. Basu A. Estimating Decision-Relevant Comparative Effects Using Instrumental Variables. *Stat Biosci*. 2011;3(1):6-27. doi:http://dx.doi.org/10.1007/s12561-011-9033-6.

93. Beck CA, Penrod J, Gyorkos TW, Shapiro S, Pilote L. Does Aggressive Care Following Acute Myocardial Infarction Reduce Mortality? Analysis with Instrumental Variables to Compare Effectiveness in Canadian and United States Patient Populations. *Health Serv Res*. 2003;38(6):1423-1440. internal-pdf://133.21.114.118/MI_mort_IV-Beck-HSR03.pdf.

94. Groenwold RH, Hak E, Klungel OH, Hoes AW. Instrumental variables in influenza vaccination studies: mission impossible?! *Value Heal*. 2010;13(1):132-137. doi:http://dx.doi.org/10.1111/j.1524-4733.2009.00584.x.

95. Hirth RA, Grabowski DC, Feng Z, Rahman M, Mor V. Effect of nursing home ownership on hospitalization of long-stay residents: An instrumental variables approach. *Int J Health Care Finance Econ*. 2014;14(1):1-18. doi:http://dx.doi.org/10.1007/s10754-013-9136-3.

96. Hollingsworth JM, Norton EC, Kaufman SR, Smith RM, Wolf Jr JS, Hollenbeck

BK. Medical expulsive therapy versus early endoscopic stone removal for acute renal colic: An instrumental variable analysis. *J Urol*. 2013;190(3):882-887. doi:http://dx.doi.org/10.1016/j.juro.2013.03.040.

97. Johnston KM, Gustafson P, Levy AR, Grootendorst P. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Stat Med*. 2008;27(9):1539-1556. doi:10.1002/sim.3036.

98. Kahn JM, Werner RM, David G, Ten Have TR, Benson NM, Asch DA. Effectiveness of long-term acute care hospitalization in elderly patients with chronic critical illness. *Med Care*. 2013;51(1):4-10. doi:http://dx.doi.org/10.1097/MLR.0b013e31826528a7.

99. Kim D, Leigh JP. Estimating the effects of wages on obesity. *J Occup Environ Med*. 2010;52(5):495-500. doi:http://dx.doi.org/10.1097/JOM.0b013e3181dbc867.

100. Lei X, Lin W. The new cooperative medical scheme in rural China: Does more coverage mean more service and better health? *Health Econ*. 2009;18(suppl. 2):S25-S46. doi:http://dx.doi.org/10.1002/hec.1501.

101. Lin MJ, Liu JT. Do lower birth weight babies have lower grades? Twin fixed effect and instrumental variable method evidence from Taiwan. *Soc Sci Med*. 2009;68(10):1780-1787. doi:http://dx.doi.org/10.1016/j.socscimed.2009.02.031.

102. Linden A, Adams JL. Evaluating disease management programme effectiveness: An introduction to instrumental variables. *J Eval Clin Pract*. 2006;12(2):148-154. doi:http://dx.doi.org/10.1111/j.1365-2753.2006.00615.x.

103. Norton EC, Lindrooth RC, Ennett ST. Controlling for the endogeneity of peer substance use on adolescent alcohol and tobacco use. *Health Econ*. 1998;7(5):439-453. doi:http://dx.doi.org/10.1002/%28SICI%291099-1050%28199808%297:5%3C439::AID-HEC362%3E3.0.CO;2-9.

104. Cawley J, Meyerhoefer C. The medical care costs of obesity: An instrumental variables approach. *J Health Econ*. 2012;31:219-230.

doi:http://dx.doi.org/10.1016/j.jhealeco.2011.10.003.

105. O'Donnell HC, Colman G, Trachtman RA, Velazco N, Racine AD. Impact of newborn follow-up visit timing on subsequent ED visits and hospital readmissions: AN instrumental variable analysis. *Acad Pediatr*. 2014;14(1):84-91. doi:http://dx.doi.org/10.1016/j.acap.2013.09.010.

106. O'Malley AJ, Frank RG, Normand SLT. Estimating cost-offsets of new medications: Use of new antipsychotics and mental health costs for schizophrenia. *Stat Med*. 2011;30(16):1971-1988. doi:http://dx.doi.org/10.1002/sim.4245.

107. Pilote L, Beck CA, Eisenberg MJ, et al. Comparing invasive and noninvasive management strategies for acute myocardial infarction using administrative databases. *Am Heart J*. 2008;155(1):42-48. doi:http://dx.doi.org/10.1016/j.ahj.2007.09.016.

108. Pirracchio R, Sprung C, Payen D, Chevret S. Benefits of ICU admission in critically ill patients: whether instrumental variable methods or propensity scores should be used. *BMC Med Res Methodol*. 2011;11(132). http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed11&AN=21936926.

109. Pracht EE, Tepas IJJ, Celso BG, Langland-Orban B, Flint L. Survival advantage associated with treatment of injury at designated trauma centers: A bivariate probit model with instrumental variables. *Med Care Res Rev*. 2007;64(1):83-97. doi:http://dx.doi.org/10.1177/1077558706296241.

110. Schmittdiel JA, Karter AJ, Dyer W, et al. The comparative effectiveness of mail order pharmacy use vs. local pharmacy use on LDL-C control in new statin users. *J Gen Intern Med*. 2011;26(12):1396-1402. doi:http://dx.doi.org/10.1007/s11606-011-1805-7.

111. Selden TM, Hudson JL. Access to care and utilization among children: Estimating the effects of public and private coverage. *Med Care*. 2006;44(5 SUPPL.):I19-I26. doi:http://dx.doi.org/10.1097/01.mlr.0000208137.46917.3b.

112. Slade EP, McCarthy JF, Valenstein M, Visnic S, Dixon LB. Cost savings from

assertive community treatment services in an era of declining psychiatric inpatient use. *Health Serv Res*. 2013;48(1):195-217. doi:http://dx.doi.org/10.1111/j.1475-6773.2012.01420.x.

113. Slade EP, Wissow LS, Davis M, Abrams MT, Dixon LB. Medicaid lapses and low-income young adults' receipt of outpatient mental health care after an inpatient stay. *Psychiatr Serv*. 2014;65(4):454-460. doi:http://dx.doi.org/10.1176/appi.ps.201200375.

114. Tsai AC, Votruba M, Bridges JFP, Cebul RD. Overcoming bias in estimating the volume-outcome relationship. *Health Serv Res*. 2006;41(1):252-264. doi:http://dx.doi.org/10.1111/j.1475-6773.2005.00461.x.

115. Chen LF, Chen HP, Huang YS, Huang KY, Chou P, Lee CC. Pneumococcal Pneumonia and the Risk of Stroke: A Population-Based Follow-Up Study. *PLoS One*. 2012;7(12):e51452. doi:http://dx.doi.org/10.1371/journal.pone.0051452.

116. Wehby GL, Ullrich F, Xie Y. Very low birth weight hospital volume and mortality: An instrumental variables approach. *Med Care*. 2012;50(8):714-721. doi:http://dx.doi.org/10.1097/MLR.0b013e31824e32cf.

117. Zeliadt SB, Loggers ET, Slatore CG, et al. Preoperative PET and the reduction of unnecessary surgery among newly diagnosed lung cancer patients in a community setting. *J Nucl Med*. 2014;55(3):379-385. doi:http://dx.doi.org/10.2967/jnumed.113.124230.

118. Edwards ST, Prentice JC, Simon SR, Pizer SD. Home-Based Primary Care and the risk of ambulatory care-sensitive condition hospitalization among older veterans with diabetes mellitus. *JAMA Intern Med*. 2014;174(11):1796-1803. doi:http://dx.doi.org/10.1001/jamainternmed.2014.4327.

119. Federspiel JJ, Stearns SC, Sheridan BC, et al. Evaluating the effectiveness of a rapidly adopted cardiovascular technology with administrative data: The case of drug-eluting stents for acute coronary syndromes. *Am Heart J*. 2012;164(2):207-214. doi:http://dx.doi.org/10.1016/j.ahj.2012.05.016.

120. Frances CD, Shlipak MG, Noguchi H, Heidenreich PA, McClellan M. Does

physician specialty affect the survival of elderly patients with myocardial infarction? *Health Serv Res*. 2000;35(5):1093-1116. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed5&AN=2001006754.

121. Goldman DP, Bao Y. Effective HIV treatment and the employment of HIV+ adults. *Health Serv Res*. 2004;39(6):1691-1712. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed6&AN=2004522727; http://lib.exeter.ac.uk:4556/resserv?sid=OVID:embase&id=pmid:&id=doi:&issn=0017-9124&isbn=&volume=39&issue=6+I&spage=1691&pages=1691-1712&date=2004&title=Health+Serv.

122. Gowrisankaran G, Town RJ. Estimating the quality of care in hospitals using instrumental variables. *J Health Econ*. 1999;18(6):747-767. doi:http://dx.doi.org/10.1016/S0167-6296%2899%2900022-3.

123. Goyal N, Zubizarreta JR, Small DS, Lorch SA. Length of stay and readmission among late preterm infants: An instrumental variable approach. *Hosp Pediatr*. 2013;3(1):7-15. doi:http://dx.doi.org/10.1542/hpeds.2012-0027.

124. O'Malley AJ. Instrumental variable specifications and assumptions for longitudinal analysis of mental health cost offsets. *Heal Serv Outcomes Res Methodol*. 2012;12(4):254-272. doi:http://dx.doi.org/10.1007/s10742-012-0097-7.

125. Richardson DB, Laurier D, Schubauer-Berigan MK, Tchetgen ET, Cole SR. Assessment and indirect adjustment for confounding by smoking in cohort studies using relative hazards models. *Am J Epidemiol*. 2014;180(9):933-940. doi:http://dx.doi.org/10.1093/aje/kwu211.

126. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative Controls. *Epidemiology*. 2010;21(3):383-388. doi:10.1097/EDE.0b013e3181d61eeb.

127. Abrahamowicz M, Beauchamp ME, Ionescu-Ittu R, Delaney JAC, Pilote L. Reducing the variance of the prescribing preference-based instrumental variable estimates of the treatment effect. *Am J Epidemiol*. 2011;174(4):494-502. doi:http://dx.doi.org/10.1093/aje/kwr057.

128. De Ridder A, De Graeve D. Can we account for selection bias? A comparison between bare metal and drug-eluting stents. *Value Heal*. 2011;14(1):3-14. doi:http://dx.doi.org/10.1016/j.jval.2010.10.014.

129. Fang G, Brooks JM, Chrischilles EA. Comparison of instrumental variable analysis using a new instrument with risk adjustment methods to reduce confounding by indication. *Am J Epidemiol*. 2012;175(11):1142-1151. doi:http://dx.doi.org/10.1093/aje/kwr448.

130. Figueroa R, Harman J, Engberg J. Use of Claims Data to Examine the Impact of Length of Inpatient Psychiatric Stay on Readmission Rate. *Psychiatr Serv*. 2004;55(5):560-565. doi:http://dx.doi.org/10.1176/appi.ps.55.5.560.

131. Guo J, Konetzka RT, Manning WG. The causal effects of home care use on institutional long-term care utilization and expenditures. *Heal Econ (United Kingdom)*. 2015;24(S1):4-17. doi:http://dx.doi.org/10.1002/hec.3155.

132. Hadley J, Polsky D, Mandelblatt JS, et al. An exploratory instrumental variable analysis of the outcomes of localized breast cancer treatments in a medicare population. *Health Econ*. 2003;12(3):171-186. doi:http://dx.doi.org/10.1002/hec.710.

133. Hay JW, Lawler E, Yucel K, et al. Cost impact of diagnostic imaging for lower extremity peripheral vascular occlusive disease. *Value Heal*. 2009;12(2):262-266. doi:http://dx.doi.org/10.1111/j.1524-4733.2008.00438.x.

134. Huesch MD. External adjustment sensitivity analysis for unmeasured confounding: An application to coronary stent outcomes, Pennsylvania 2004-2008. *Health Serv Res*. 2013;48(3):1191-1214. doi:http://dx.doi.org/10.1111/1475-6773.12013.

135. Huybrechts KF, Brookhart MA, Rothman KJ, et al. Comparison of different approaches to confounding adjustment in a study on the association of antipsychotic medication with mortality in older nursing home patients. *Am J Epidemiol*. 2011;174(9):1089-1099. doi:http://dx.doi.org/10.1093/aje/kwr213.

136. Ionescu-Ittu R, Abrahamowicz M, Pilote L. Treatment effect estimates varied depending on the definition of the provider prescribing preference-based

instrumental variables. *J Clin Epidemiol*. 2012;65(2):155-162. doi:http://dx.doi.org/10.1016/j.jclinepi.2011.06.012.

137. Kivimaki M, Vahtera J, Kawachi I, et al. Psychosocial work environment as a risk factor for absence with a psychiatric diagnosis: An instrumental-variables analysis. *Am J Epidemiol*. 2010;172(2):167-172. doi:http://dx.doi.org/10.1093/aje/kwq094.

138. Ahern TP, Pedersen L, Svaerke C, Rothman KJ, Sorensen HT, Lash TL. The association between vitamin K antagonist therapy and site-specific cancer incidence estimated by using heart valve replacement as an instrumental variable. *Am J Epidemiol*. 2011;174(12):1382-1390. doi:http://dx.doi.org/10.1093/aje/kwr268.

139. Kramer A, Jager KJ, Fogarty DG, et al. Association between pre-transplant dialysis modality and patient and graft survival after kidney transplantation. *Nephrol Dial Transplant*. 2012;27(12):4473-4480. doi:http://dx.doi.org/10.1093/ndt/gfs450.

140. Kuo YF, Montie JE, Shahinian VB. Reducing bias in the assessment of treatment effectiveness: Androgen deprivation therapy for prostate cancer. *Med Care*. 2012;50(5):374-380. doi:http://dx.doi.org/10.1097/MLR.0b013e318245a086.

141. Lakdawalla DN, Mascarenhas M, Jena AB, et al. Impact of oral nutrition supplements on hospital outcomes in pediatric patients. *J Parenter Enter Nutr*. 2014;38(6):42S-49S. doi:http://dx.doi.org/10.1177/0148607114549769.

142. MacKenzie TA, Tosteson TD, Morden NE, Stukel TA, O'Malley AJ. Using instrumental variables to estimate a Cox's proportional hazards regression subject to additive confounding. *Heal Serv Outcomes Res Methodol*. 2014;14(1-2):54-68. doi:http://dx.doi.org/10.1007/s10742-014-0117-x.

143. Margolis DJ, Gupta J, Hoffstad O, et al. Lack of effectiveness of hyperbaric oxygen therapy for the treatment of diabetic foot ulcer and the prevention of amputation a cohort study. *Diabetes Care*. 2013;36(7):1961-1966. doi:http://dx.doi.org/10.2337/dc12-2160.

144. Newman TB, Vittinghoff E, McCulloch CE. Efficacy of phototherapy for newborns with hyperbilirubinemia: a cautionary example of an instrumental variable analysis. *Med Decis Making*. 2012;32(1):83-92. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed10&AN=21859678.

145. Parmar AD, Sheffield KM, Han Y, et al. Evaluating comparative effectiveness with observational data: Endoscopic ultrasound and survival in pancreatic cancer. *Cancer*. 2013;119(21):3861-3869. doi:http://dx.doi.org/10.1002/cncr.28295.

146. Pisoni RL, Arrington CJ, Albert JM, et al. Facility Hemodialysis Vascular Access Use and Mortality in Countries Participating in DOPPS: An Instrumental Variable Analysis. *Am J Kidney Dis*. 2009;53(3):475-491. doi:http://dx.doi.org/10.1053/j.ajkd.2008.10.043.

147. Prentice JC, Conlin PR, Gellad WF, Edelman D, Lee TA, Pizer SD. Capitalizing on prescribing pattern variation to compare medications for type 2 diabetes. *Value Heal*. 2014;17(8):854-862. doi:http://dx.doi.org/10.1016/j.jval.2014.08.2674.

148. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol*. 2009;62(12):1233-1241. doi:http://dx.doi.org/10.1016/j.jclinepi.2008.12.006.

149. An J, Nichol MB. Multiple medication adherence and its effect on clinical outcomes among patients with comorbid type 2 diabetes and hypertension. *Med Care*. 2013;51(10):879-887. doi:http://dx.doi.org/10.1097/MLR.0b013e31829fa8ed.

150. Rosenthal MB, Li Z, Robertson AD, Milstein A. Impact of financial incentives for prenatal care on birth outcomes and spending. *Health Serv Res*. 2009;44(5 Part 1):1465-1479. doi:http://dx.doi.org/10.1111/j.1475-6773.2009.00996.x.

151. Sheffield KM, Riall TS, Han Y, Kuo YF, Townsend, C. M. J, Goodwin JS. Association between cholecystectomy with vs without intraoperative

cholangiography and risk of common duct injury. *Jama*. 2013;310(8):812-820. doi:http://dx.doi.org/10.1001/jama.2013.276205.

152. Steingrub JS, Lagu T, Rothberg MB, Nathanson BH, Raghunathan K, Lindenauer PK. Treatment with neuromuscular blocking agents and the risk of in-hospital mortality among mechanically ventilated patients with severe sepsis. *Crit Care Med*. 2014;42(1):90-96. doi:http://dx.doi.org/10.1097/CCM.0b013e31829eb7c9.

153. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. *JAMA*. 2007;297(3):278-285. doi:10.1001/jama.297.3.278.

154. Tagami T, Matsui H, Horiguchi H, Fushimi K, Yasunaga H. Antithrombin and mortality in severe pneumonia patients with sepsis-associated disseminated intravascular coagulation: An observational nationwide study. *J Thromb Haemost*. 2014;12(9):1470-1479. doi:http://dx.doi.org/10.1111/jth.12643.

155. VanDyke RD, McPhail GL, Huang B, et al. Inhaled tobramycin effectively reduces FEV1 decline in cystic fibrosis an instrumental variables analysis. *Ann Am Thorac Soc*. 2013;10(3):205-212. doi:http://dx.doi.org/10.1513/AnnalsATS.201209-082OC.

156. Wong K, Campitelli MA, Stukel TA, Kwong JC. Estimating influenza vaccine effectiveness in community-dwelling elderly patients using the instrumental variable analysis method. *Arch Intern Med*. 2012;172(6):484-491. doi:10.1001/archinternmed.2011.2038.

157. Bekelman JE, Mitra N, Handorf EA, et al. Effectiveness of androgen-deprivation therapy and radiotherapy for older men with locally advanced prostate cancer. *J Clin Oncol*. 2015;33(7):716-722. doi:http://dx.doi.org/10.1200/JCO.2014.57.2743.

158. Bhowmik D, Aparasu RR, Rajan SS, Sherer JT, Ochoa-Perez M, Chen H. Risk of manic switch associated with antidepressant therapy in pediatric bipolar depression. *J Child Adolesc Psychopharmacol*. 2014;24(10):551-561.

doi:http://dx.doi.org/10.1089/cap.2014.0028.

159. Brooks JM, Tang Y, Chapman CG, Cook EA, Chrischilles EA. What is the effect of area size when using local area practice style as an instrument? *J Clin Epidemiol*. 2013;66(8 SUPPL.8):S69-S83. doi:http://dx.doi.org/10.1016/j.jclinepi.2013.04.008.

160. Cai B, Hennessy S, Flory JH, Sha D, Ten Have TR, Small DS. Simulation study of instrumental variable approaches with an application to a study of the antidiabetic effect of bezafibrate. *Pharmacoepidemiol Drug Saf*. 2012;21(SUPPL.2):114-120. doi:http://dx.doi.org/10.1002/pds.3252.

161. Chen H, Mehta S, Aparasu R, Patel A, Ochoa-Perez M. Comparative effectiveness of monotherapy with mood stabilizers versus second generation (atypical) antipsychotics for the treatment of bipolar disorder in children and adolescents. *Pharmacoepidemiol Drug Saf*. 2014;23(3):299-308. doi:http://dx.doi.org/10.1002/pds.3568.

162. Chuang CM, Chou YJ, Yen MS, et al. The role of secondary cytoreductive surgery in patients with recurrent epithelial ovarian, tubal, and peritoneal cancers: A comparative effectiveness analysis. *Oncologist*. 2012;17(6):847-855. doi:http://dx.doi.org/10.1634/theoncologist.2011-0373.

163. Bryson WC, McConnell J, Krothuis T, McCarty D. Extended-release naltrexone for alcohol dependence: persistence and healthcare costs and utilization. *Am J Manag Care*. 2011;17 Suppl 8:S222-234. internal-pdf://72.27.23.57/alcohol_depend-Bryson-AmJManCare11.pdf internal-pdf://4105231440/Bryson2011-suppl_appdx.docx.

164. Cheng L, Liu H, Zhang Y, Shen K, Zeng Y. The impact of health insurance on health outcomes and spending of the elderly: Evidence from china's new cooperative medical scheme. *Heal Econ (United Kingdom)*. 2015;24(6):672-691. doi:http://dx.doi.org/10.1002/hec.3053.

165. Menon J, Paulet M, Thomas IJ. Wellness coaching and health-related quality of life: A case-control difference-in-differences analysis. *J Occup Environ Med*. 2012;54(10):1259-1267. doi:http://dx.doi.org/10.1097/JOM.0b013e31825a2594.

166. Moran JR, Short PF, Hollenbeak CS. Long-term employment effects of surviving cancer. *J Health Econ*. 2011;30(3):505-514. doi:http://dx.doi.org/10.1016/j.jhealeco.2011.02.001.

167. Osborne NH, Nicholas LH, Ryan AM., Humma JR, Dimick JB. Association of hospital participation in a quality reporting program with surgical outcomes and expenditures for medicare beneficiaries. *JAMA - J Am Med Assoc*. 2015;313(5):496-504. doi:http://dx.doi.org/10.1001/jama.2015.25.

168. Rajaram R, Chung JW, Jones AT, et al. Association of the 2011 ACGME resident duty hour reform with general surgery patient outcomes and with resident examination performance. *JAMA - J Am Med Assoc*. 2014;312(22):2374-2384. doi:http://dx.doi.org/10.1001/jama.2014.15277.

169. Reid RO, Ashwood JS, Friedberg MW, Weber ES, Setodji CM, Mehrotra A. Retail clinic visits and receipt of primary care. *J Gen Intern Med*. 2013;28(4):504-512. doi:http://dx.doi.org/10.1007/s11606-012-2243-x.

170. Sadhu AR, Ang AC, Ingram-Drake LA, Martinez DS, Hsueh WA, Ettner SL. Economic benefits of intensive insulin therapy in critically Ill patients: The targeted insulin therapy to improve hospital outcomes (TRIUMPH) project. *Diabetes Care*. 2008;31(8):1556-1561. doi:http://dx.doi.org/10.2337/dc07-2456.

171. Sarkar U, Lyles CR, Parker MM, et al. Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system. *Med Care*. 2014;52(3):194-201. doi:http://dx.doi.org/10.1097/MLR.0000000000000069.

172. Watt C, Abuya T, Warren CE, Obare F, Kanya L, Bellows B. Can reproductive health voucher programs improve quality of postnatal care? A quasi-experimental evaluation of Kenya ' s Safe Motherhood voucher scheme. *PLoS One*. 2015;10(4). doi:http://dx.doi.org/10.1371/journal.pone.0122828 April.

173. De Preux LB. Anticipatory ex ante moral hazard and the effect of medicare on prevention. *Health Econ*. 2011;20(9):1056-1072. doi:http://dx.doi.org/10.1002/hec.1778.

174. Gebel M, Voßemer J. The impact of employment transitions on health in Germany. A difference-in-differences propensity score matching approach. *Soc Sci Med*. 2014;108:128-136. doi:10.1016/j.socscimed.2014.02.039.

175. Goetzel RZ, Roemer EC, Pei X, et al. Second-year results of an obesity prevention program at the dow chemical company. *J Occup Environ Med*. 2010;52(3):291-302. doi:http://dx.doi.org/10.1097/JOM.0b013e3181d46f0b.

176. Higgins S, Chawla R, Colombo C, Snyder R, Nigam S. Medical homes and cost and utilization among high-risk patients. *Am J Manag Care*. 2014;20(3):e61-e71. internal-pdf://67.10.22.98/medhomes_cost_highrisk-Higgins-AJMC14.pdf.

177. Kausto J, Viikari-Juntura E, Virta LJ, Gould R, Koskinen A, Solovieva S. Effectiveness of new legislation on partial sickness benefit on work participation: a quasi-experiment in Finland. *BMJ Open*. 2014;4(12):e006685. doi:http://dx.doi.org/10.1136/bmjopen-2014-006685.

178. Kelly Y, Kelly J, Sacker A. Changes in bedtime schedules and behavioral difficulties in 7 year old children. *Pediatrics*. 2013;132(5):e1184-e1193. doi:http://dx.doi.org/10.1542/peds.2013-1906.

179. Lin WC, Chien HL, Willis G, et al. The effect of a telephone-based health coaching disease management program on medicaid members with chronic conditions. *Med Care*. 2012;50(1):91-98. doi:http://dx.doi.org/10.1097/MLR.0b013e31822dcedf.

180. Lyon SM, Wunsch H, Asch DA, Carr BG, Kahn JM, Cooke CR. Use of intensive care services and associated hospital mortality after massachusetts healthcare reform. *Crit Care Med*. 2014;42(4):763-770. doi:http://dx.doi.org/10.1097/CCM.0000000000000044.

181. Domino ME, Norton EC, Morrissey JP, Thakur N. Cost shifting to jails after a change to managed mental health care. *Health Serv Res*. 2004;39(5):1379-1401. doi:http://dx.doi.org/10.1111/j.1475-6773.2004.00295.x.

182. Hodgkin D, Parks Thomas C, Simoni-Wastila L, Ritter GA, Lee S. The effect of a three-tier formulary on antidepressant utilization and expenditures. *J Ment*

*Heal Policy Econ*. 2008;11(2):67-77. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med5&AN=18509214; http://lib.exeter.ac.uk:4556/resserv?sid=OVID:medline&id=pmid:18509214&id=doi:&issn=1091-4358&isbn=&volume=11&issue=2&spage=67&pages=67-77&date=2008&title=The+Journal+o.

183. Li J, Hurley J, DeCicca P, Buckley G. Physician response to pay-for-performance: evidence from a natural experiment. *Health Econ*. 2014;23(8):962-978. doi:http://dx.doi.org/10.1002/hec.2971.

184. Yoon J, Bernell SL. The role of adverse physical health events on the utilization of mental health services. *Health Serv Res*. 2013;48(1):175-194. doi:http://dx.doi.org/10.1111/j.1475-6773.2012.01442.x.

185. Wagstaff A, Moreno-Serra R. Europe and central Asia's great post-communist social health insurance experiment: Aggregate impacts on health sector outcomes. *J Health Econ*. 2009;28(2):322-340. doi:10.1016/j.jhealeco.2008.10.011.

186. Bell B, Blundell R, Reenen J Van. Getting the Unemployed Back to Work: The Role of Targeted Wage Subsidies. *Int Tax Public Financ*. 1999;6(3):339-360. doi:10.1023/A:1008787013977.

187. Tannen RL, Weiner MG, Xie D. Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: Further empiric validation of the "prior event rate ratio" to adjust for unmeasured confounding by indication. *Pharmacoepidemiol Drug Saf*. 2008;17(7):671-685. doi:http://dx.doi.org/10.1002/pds.1584.

188. Tannen R, Weiner M, Xie D. Use of Primary Care Electronic Medical Record Database in Drug Efficacy Research on Cardiovascular Outcomes: Comparison of Database and Randomised. *BMJ Br Med J*. 2009;338(b91):1-9. http://www.jstor.org/stable/10.2307/20512072. Accessed January 9, 2014.

189. Tannen R, Wang X, Yu M, Weiner MG, Xie D. A new "Comparative Effectiveness" assessment strategy using the THIN database: Comparison of the cardiac complications of pioglitazone and rosiglitazone.

*Pharmacoepidemiol Drug Saf*. 2013;22(1):86-97. doi:http://dx.doi.org/10.1002/pds.3360.

190. Brophy S, Jones KH, Rahman MA, et al. Incidence of Campylobacter and Salmonella Infections Following First Prescription for PPI: A Cohort Study Using Routine Data. *Am J Gastroenterol*. 2013;108:1094–1100. doi:10.1038/ajg.2013.30.

191. Uddin MJ, Groenwold RHH, van Staa TP, et al. Performance of prior event rate ratio adjustment method in pharmacoepidemiology: a simulation study. *Pharmacoepidemiol Drug Saf*. November 2014. doi:10.1002/pds.3724.

192. Lin NX, Henley WE. Prior event rate ratio adjustment for hidden confounding in observational studies of treatment effectiveness: a pairwise Cox likelihood approach. *Stat Med*. 2016;35(28):5149-5169. doi:10.1002/sim.7051.

193. Albouy V, Lequien L. Does compulsory education lower mortality? *J Health Econ*. 2009;28(1):155-168. doi:http://dx.doi.org/10.1016/j.jhealeco.2008.09.003.

194. Swaminathan S, Mor V, Mehrotra R, Trivedi AN. Effect of medicare dialysis payment reform on use of erythropoiesis stimulating agents. *Health Serv Res*. 2015;50(3):790-808. doi:http://dx.doi.org/10.1111/1475-6773.12252.

195. Zuckerman IH, Lee E, Wutoh AK, Xue Z, Stuart B. Application of regression-discontinuity analysis in pharmaceutical health services research. *Health Serv Res*. 2006;41(2):550-563. doi:10.1111/j.1475-6773.2005.00487.x.

196. Wagner TH, Jimison HB. Computerized health information and the demand for medical care. *Value Heal*. 2003;6(1):29-39. doi:http://dx.doi.org/10.1046/j.1524-4733.2003.00155.x.

197. Fortney JC, Steffick DE, Burgess Jr JF, Maciejewski ML, Petersen LA. Are primary care services a substitute or complement for specialty and inpatient services? *Health Serv Res*. 2005;40(5):1422-1442. doi:http://dx.doi.org/10.1111/j.1475-6773.2005.00424.x.

198. Chung S, Domino ME, Stearns SC. The effect of retirement on weight. *journals Gerontol*. 2009;Series B,(5):656-665.

http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed9&AN=19357073; http://lib.exeter.ac.uk:4556/resserv?sid=OVID:embase&id=pmid:19357073&id=doi:&issn=1758-5368&isbn=&volume=64&issue=5&spage=656&pages=656-665&date=2009&title=The+journa.

199. Hay J, Jhaveri M, Tangirala M, Kaliner M. Cost and resource utilization comparisons of second-generation antihistamines vs. montelukast for allergic rhinitis treatment. *Allergy Asthma Proc*. 2009;30(6):634-642. doi:http://dx.doi.org/10.2500/aap.2009.30.3293.

200. Kawatkar AA, Hay JW, Stohl W, Nichol MB. Incremental expenditure of biologic disease modifying antirheumatic treatment using instrumental variables in panel data. *Heal Econ (United Kingdom)*. 2013;22(7):807-823. doi:http://dx.doi.org/10.1002/hec.2855.

201. Piernas C, Ng SW, Mendez MA, Gordon-Larsen P, Popkin BM. A dynamic panel model of the associations of sweetened beverage purchases with dietary quality and food-purchasing patterns. *Am J Epidemiol*. 2015;181:661-671. doi:http://dx.doi.org/10.1093/aje/kwu317.

202. Arellano M, Bond S. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Rev Econ Stud*. 1991;58(2):277. doi:10.2307/2297968.

203. Arellano M, Bover O. Another look at the instrumental variable estimation of error-components models. *J Econom*. 1995;68(1):29-51. doi:10.1016/0304-4076(94)01642-D.

204. Blundell R, Bond S. Initial conditions and moment restrictions in dynamic panel data models. *J Econom*. 1998;87(1):115-143. doi:10.1016/S0304-4076(98)00009-8.

205. MacKenzie TA, Tosteson TD, Morden NE, Stukel TA, O'Malley AJ. Using instrumental variables to estimate a Cox's proportional hazards regression subject to additive confounding. *Heal Serv Outcomes Res Methodol*. 2014;14(1-2):54-68. doi:10.1007/s10742-014-0117-x.

206. Jackson JW, Swanson SA. Toward a clearer portrayal of confounding bias in instrumental variable applications. *Epidemiology*. 2015;26(4):498-504. doi:10.1097/EDE.0000000000000287.

207. Davies NM. An even clearer portrait of bias in observational studies? *Epidemiology*. 2015;26(4):505-508. doi:10.1097/EDE.0000000000000302.

208. Uddin MJ, Groenwold RHH, Ali MS, et al. Methods to control for unmeasured confounding in pharmacoepidemiology: an overview. *Int J Clin Pharm*. April 2016. doi:10.1007/s11096-016-0299-0.

209. Alemayehu D, Alvir JMJ, Jones B, Willke RJ. Statistical issues with the analysis of nonrandomized studies in comparative effectiveness research. *J Manag Care Pharm*. 2011;17(9 Suppl A):S22-6. http://www.ncbi.nlm.nih.gov/pubmed/22074671. Accessed July 18, 2014.

210. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM. *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*. Agency for Healthcare Research and Quality (US); 2013. http://www.ncbi.nlm.nih.gov/pubmed/23469377. Accessed October 25, 2016.

211. Freemantle N, Marston L, Walters K, Wood J, Reynolds MR, Petersen I. Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. *Bmj*. 2013;347(nov11 3):f6409-f6409. doi:10.1136/bmj.f6409.

212. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res*. 2011;46(3):399-424. doi:10.1080/00273171.2011.568786.

213. Bosco JL, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol*. 2010;63(1):64-74. doi:http://dx.doi.org/10.1016/j.jclinepi.2009.03.001.

214. Moscoe E, Bor J, Bärnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current

and best practice. *J Clin Epidemiol*. 2015;68(2):122-133. doi:10.1016/j.jclinepi.2014.06.021.

215. Nordmann S, Biard L, Ravaud P, et al. Case-Only Designs in Pharmacoepidemiology: A Systematic Review. Little J, ed. *PLoS One*. 2012;7(11):e49444. doi:10.1371/journal.pone.0049444.

216. Cornfield J, Haenszel W. Smoking and lung cancer: recent evidence and a discussion of some questions. *J Nat Cancer …*. 1959. http://www.epidemiology.ch/history/PDF bg/Cornfield J et al 2009 smoking and lung cancer - recent evidence.pdf. Accessed February 6, 2014.

217. Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol*. 1996;25(6):1107-1116. doi:10.1093/ije/25.6.1107.

218. Margolis DJ, Berlin JA, Strom BL. A comparison of sensitivity analyses of the effect of wound duration on wound healing. *J Clin Epidemiol*. 1999;52(2):123-128. doi:http://dx.doi.org/10.1016/S0895-4356%2898%2900150-4.

219. Rosenbaum PR. *Observational Studies*. New York, NY: Springer New York; 1995. doi:10.1007/978-1-4757-2443-1.

220. Cabral R. R. MD. L. Use of sensitivity analysis to assess the effects on anti-hepatitis A virus antibodies of access to household water supply. *Epidemiol Infect*. 2008;136(3):334-340. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med5&AN=17445315; http://lib.exeter.ac.uk:4556/resserv?sid=OVID:medline&id=pmid:17445315&id=doi:&issn=0950-2688&isbn=&volume=136&issue=3&spage=334&pages=334-40&date=2008&title=Epidemiolo.

221. Arah OA, Chiba Y, Greenland S. Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. *Ann Epidemiol*. 2008;18(8):637-646. doi:10.1016/j.annepidem.2008.04.003.

222. Vanderweele TJ, Mukherjee B, Chen J. Sensitivity analysis for interactions under unmeasured confounding. *Stat Med*. 2012;31(22):2552-2564. doi:http://dx.doi.org/10.1002/sim.4354.

223. MacLehose RF, Kaufman S, Kaufman JS, Poole C. Bounding causal effects under uncontrolled confounding using counterfactuals. *Epidemiology*. 2005;16:548-555. doi:http://dx.doi.org/10.1097/01.ede.0000166500.23446.53.

224. Lin DY, Psaty BM, Kronmal RA. Assessing the Sensitivity of Regression Results to Unmeasured Confounders in Observational Studies. *Biometrics*. 1998;54(3):948. doi:10.2307/2533848.

225. Lin NX, Logan S, Henley WE. Bias and sensitivity analysis when estimating treatment effects from the cox model with omitted covariates. *Biometrics*. 2013;69(4):850-860. doi:10.1111/biom.12096.

226. Steenland K, Greenland S. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *Am J Epidemiol*. 2004;160(4):384-392. doi:10.1093/aje/kwh211.

227. De Vocht F, Kromhout H, Ferro G, Boffetta P, Burstyn I. Bayesian modelling of lung cancer risk and bitumen fume exposure adjusted for unmeasured confounding by smoking. *Occup Environ Med*. 2009;66(8):502-508. doi:http://dx.doi.org/10.1136/oem.2008.042606.

228. Corrao G, Nicotra F, Parodi A, et al. External adjustment for unmeasured confounders improved drug-outcome association estimates based on health care utilization data. *J Clin Epidemiol*. 2012;65(11):1190-1199. doi:http://dx.doi.org/10.1016/j.jclinepi.2012.03.014.

229. McCandless LC, Gustafson P, Levy A. Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Stat Med*. 2007;26(11):2331-2347. doi:http://dx.doi.org/10.1002/sim.2711.

230. McCandless LC, Gustafson P, Levy AR. A sensitivity analysis using information about measured confounders yielded improved uncertainty assessments for unmeasured confounding. *J Clin Epidemiol*. 2008;61(3):247-255. doi:http://dx.doi.org/10.1016/j.jclinepi.2007.05.006.

231. Gustafson P, McCandless LC, Levy AR, Richardson S. Simplified Bayesian Sensitivity Analysis for Mismeasured and Unobserved Confounders. *Biometrics*. 2010;66:1129-1137. doi:http://dx.doi.org/10.1111/j.1541-

0420.2009.01377.x.

232. Brunelli SM, Gagne JJ, Huybrechts KF, et al. Estimation using all available covariate information versus a fixed look-back window for dichotomous covariates. *Pharmacoepidemiol Drug Saf*. 2013;22:542-550. doi:http://dx.doi.org/10.1002/pds.3434.

233. Sturmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am J Epidemiol*. 2005;162:279-289.

234. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf*. 2006;15:291-303. doi:http://dx.doi.org/10.1002/pds.1200.

235. Li L, Shen C, Wu AC, Li X. Propensity score-based sensitivity analysis method for uncontrolled confounding. *Am J Epidemiol*. 2011;174(3):345-353. doi:http://dx.doi.org/10.1093/aje/kwr096.

236. Robins J, Rotnitzky A, Scharfstein D. *Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models*.; 2000. http://link.springer.com/chapter/10.1007/978-1-4612-1284-3_1. Accessed January 13, 2015.

237. Brumback BA, Hernán MA, Haneuse SJPA, Robins JM. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Stat Med*. 2004;23(5):749-767. doi:10.1002/sim.1657.

238. Ko  Hogan, Joseph W. and Mayer, Kenneth H H. Estimating causal treatment effects from longitudinal HIV natural history studies using marginal structural models . *Biometrics*. 2003;59:152-162.

239. Robins JM. Association, Causation, And Marginal Structural Models. *Synthese*. 1999;121(1-2):151-179. doi:10.1023/A:1005285815569.

240. Hernán MA, Brumback B, Robins JM. Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments. *J Am Stat Assoc*. 2001;96(454):440-448.

http://amstat.tandfonline.com/doi/abs/10.1198/016214501753168154. Accessed April 6, 2016.

241. Brumback BA, Hernan MA, Haneuse SJ, Robins JM. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Stat Med*. 2004;23(5):749-767. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med5&AN=14981673.

242. Weiner MG, Xie D, Tannen RL. Replication of the Scandinavian Simvastatin Survival Study using a primary care medical record database prompted exploration of a new method to address unmeasured confounding. *Pharmacoepidemiol Drug Saf*. 2008;17(7):661-670. doi:10.1002/pds.1585.

243. Tannen RL, Weiner MG, Xie D. Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: further empiric validation of the "prior event rate ratio" to adjust for unmeasured confounding by indication. *Pharmacoepidemiol Drug Saf*. 2008;17(7):671-685. doi:10.1002/pds.1584.

244. Cox DR. Regression Models and Life-Tables. *J R Stat Soc Ser B*. 1972;34(2):187-220.

245. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: Comparison of database and randomised controlled trial findings. *BMJ*. 2009;338(b81):1-9. doi:http://dx.doi.org/10.1136/bmj.b81.

246. Yu M, Xie D, Wang X, Weiner MG, Tannen RL. Prior event rate ratio adjustment: numerical studies of a statistical method to address unrecognized confounding in observational studies. *Pharmacoepidemiol Drug Saf*. 2012;21(Suppl 2):60-68. doi:10.1002/pds.3235.

247. Uddin R. H. H.; Van Staa, T. P.; De Boer, A.; Belitser, S. V.; Hoes, A. W.; Roes, K. C. B.; Klungel, O. H. MJ. G. Performance of prior event rate ratio adjustment method in pharmacoepidemiology: A simulation study. *Pharmacoepidemiol Drug Saf*. 2015;24:468-477. doi:http://dx.doi.org/10.1002/pds.3724.

248. Collett D. *Modelling Survival Data in Medical Research*. Second Edi. Chapman & Hall/CRC; 2014.

249. Gallagher AM, de Vries F, van Staa T. Prior Event Rate Ratio Adjustment: A Magic Bullet or More of the Same? In: *Abstracts of the 25th International Conference on Pharmacoepidemiology & Therapeutic Risk Management. Providence, Rhode Island, USA. August 16-19, 2009.* Vol 18 Suppl 1. Rhode Island: John Wiley & Sons, Ltd; 2009:S14. doi:10.1002/pds.1806.

250. Psaty BM, Siscovick DS. Minimizing Bias Due to Confounding by Indication in Comparative Effectiveness Research. *JAMA*. 2010;304(8):897. doi:10.1001/jama.2010.1205.

251. Kissling E, Valenciano M, Falcao J, et al. "I-MOVE" towards monitoring seasonal and pandemic influenza vaccine effectiveness: lessons learnt from a pilot multi-centric case-control study in Europe, 2008-9. *Euro Surveill*. 2009;14(44):19388. doi:10.2807/ese.14.44.19388-en.

252. Foppa IM, Haber M, Ferdinands JM, Shay DK. The case test-negative design for studies of the effectiveness of influenza vaccine. *Vaccine*. 2013;31(30):3104-3109. doi:10.1016/j.vaccine.2013.04.026.

253. Darvishian M, Bijlsma MJ, Hak E, van den Heuvel ER. Effectiveness of seasonal influenza vaccine in community-dwelling elderly people: a meta-analysis of test-negative design case-control studies. *Lancet Infect Dis*. 2014;14(12):1228-1239. doi:10.1016/S1473-3099(14)70960-0.

254. van Doorn E, Darvishian M, Dijkstra F, et al. Influenza vaccine effectiveness estimates in the Dutch population from 2003 to 2014: The test-negative design case-control study with different control groups. *Vaccine*. 2017;35(21):2831-2839. doi:10.1016/j.vaccine.2017.04.012.

255. Fukushima W, Hirota Y. Basic principles of test-negative design in evaluating influenza vaccine effectiveness. *Vaccine*. 2017;35(36):4796-4800. doi:10.1016/J.VACCINE.2017.07.003.

256. Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics*. 1995;51(1):228-235.

http://www.ncbi.nlm.nih.gov/pubmed/7766778. Accessed September 30, 2014.

257. Whitaker HJ, Hocine MN, Farrington CP. The methodology of self-controlled case series studies. *Stat Methods Med Res*. 2009;18(1):7-26. doi:10.1177/0962280208092342.

258. Taylor B, Miller E, Farrington CP, et al. Autism and measles, mumps, and rubella vaccine: no epidemiological evidence for a causal association.[comment]. *Lancet*. 1999;353(9169):2026-2029. doi:10.1016/S0140-6736(99)01239-8.

259. Maclure M. The Case-Crossover Design: A Method for Studying Transient Effects on the Risk of Acute Events. *Am J Epidemiol*. 1991;133(2):144-153. doi:10.1093/oxfordjournals.aje.a115853.

260. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Model*. 1986;7(9-12):1393-1512. doi:10.1016/0270-0255(86)90088-6.

261. Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JAC. Methods for dealing with time-dependent confounding. *Stat Med*. 2013;32(9):1584-1618. doi:10.1002/sim.5686.

262. Tannen R, Weiner M, Marcus S. Simulation of the Syst-Eur randomized control trial using a primary care electronic medical record was feasible. *J Clin Epidemiol*. 2006;59(3):254-264. http://www.sciencedirect.com/science/article/pii/S0895435605003100. Accessed January 9, 2014.

263. Tannen RL, Weiner MG, Xie D, Barnhart K. A simulation using data from a primary care practice database closely replicated the women's health initiative trial. *J Clin Epidemiol*. 2007;60(7):686-695. doi:10.1016/j.jclinepi.2006.10.012.

264. Tannen RL, Weiner MG, Xie D, Barnhart K. Estrogen affects post-menopausal women differently than estrogen plus progestin replacement therapy. *Hum Reprod*. 2007;22(6):1769-1777. doi:10.1093/humrep/dem031.

265. Othman F, Crooks CJ, Card TR. Community acquired pneumonia incidence

before and after proton pump inhibitor prescription: population based study. *BMJ*. 2016;355:i5813. doi:10.1136/bmj.i5813.

266. Zirk-Sadowski J, Masoli JA, Delgado J, et al. Proton-Pump Inhibitors and Long-Term Risk of Community-Acquired Pneumonia in Older Adults. *J Am Geriatr Soc*. April 2018. doi:10.1111/jgs.15385.

267. Dennis M, Shine L, John A, et al. Risk of Adverse Outcomes for Older People with Dementia Prescribed Antipsychotic Medication: A Population Based e-Cohort Study. *Neurol Ther*. January 2017:1-21. doi:10.1007/s40120-016-0060-6.

268. Brophy S, Kennedy J, Fernandez-Gutierrez F, et al. Characteristics of Children Prescribed Antipsychotics: Analysis of Routinely Collected Data. *J Child Adolesc Psychopharmacol*. 2018;28(3):1-12. doi:10.1089/cap.2017.0003.

269. Young-Xu Y, Robertus Van Aalst V, MMahmud S, et al. Relative Vaccine Effectiveness of High-Dose versus Standard-Dose Influenza Vaccines among Veterans Health Administration Patients. *J Infect Dis*. 2018;217(11):1718-1727. doi:10.1093/infdis/jiy088/4858294.

270. Armstrong BG, Mangtani P, Fletcher A, et al. Effect of influenza vaccination on excess deaths occurring during periods of high circulation of influenza: cohort study in elderly people. *BMJ*. 2004;329(7467):660. doi:10.1136/bmj.38198.594109.AE.

271. Baxter R, Ray GT, Fireman BH. Effect of influenza vaccination on hospitalizations in persons aged 50 years and older. *Vaccine*. 2010;28(45):7267-7272. doi:10.1016/j.vaccine.2010.08.088.

272. Tannen R, Yu M. A new method to address unmeasured confounding of mortality in observational studies. *Learn Heal Syst*. 2017;1(1):e10016. doi:10.1002/lrh2.10016.

273. Santos-Gallego CG, Badimon JJ. The sum of two evils: Pneumonia and myocardial infarction: Is platelet activation the missing link? *J Am Coll Cardiol*. 2014;64(18):1926-1928. doi:10.1016/j.jacc.2014.08.023.

274. Austin PC, Grootendorst P, Normand S-LT, Anderson GM. Conditioning on the

propensity score can result in biased estimation of common measures of treatment effect: a Monte Carlo study. *Stat Med*. 2007;26(4):754-768. doi:10.1002/sim.2618.

275. Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. *Stat Med*. 2013;32(16):2837-2849. doi:10.1002/sim.5705.

276. Holt JD, Prentice RL. Survival analyses in twin studies and matched pair experiments. *Biometrika*. 1974;61(1):17-30. doi:10.1093/biomet/61.1.17.

277. R Core Team. R: A language and environment for statistical computing. 2013. http://www.r-project.org/.

278. Andersen PK, Gill RD. Cox's Regression Model for Counting Processes: A Large Sample Study. *Ann Stat*. 1982;10(4):1100-1120. http://projecteuclid.org/euclid.aos/1176345976. Accessed October 18, 2013.

279. Welte T, Torres A, Nathwani D. Clinical and economic burden of community-acquired pneumonia among adults in Europe. *Thorax*. 2012;67(1):71-79. doi:10.1136/thx.2009.129502.

280. Moberley S, Holden J, Tatham DP, Andrews RM. Vaccines for preventing pneumococcal infection in adults. *Cochrane database Syst Rev*. 2013;1:CD000422. doi:10.1002/14651858.CD000422.pub3.

281. Diao W-Q, Shen N, Yu P-X, Liu B-B, He B. Efficacy of 23-valent pneumococcal polysaccharide vaccine in preventing community-acquired pneumonia among immunocompetent adults: A systematic review and meta-analysis of randomized trials. *Vaccine*. 2016;34(13):1496-1503. doi:10.1016/j.vaccine.2016.02.023.

282. Kraicer-Melamed H, O'Donnell S, Quach C. The effectiveness of pneumococcal polysaccharide vaccine 23 (PPV23) in the general population of 50 years of age and older: A systematic review and meta-analysis. *Vaccine*. 2016;34(13):1540-1550. doi:10.1016/j.vaccine.2016.02.024.

283. Kraicer-Melamed H, O'Donnell S, Quach C. Corrigendum to "The effectiveness of pneumococcal polysaccharide vaccine 23 (PPV23) in the general population

of 50years of age and older: A systematic review and meta-analysis" [Vaccine 34 (2016) 1540–1550]. *Vaccine*. 2016;34(34):4083-4084. doi:10.1016/j.vaccine.2016.06.045.

284. Schiffner-Rohe J, Witt A, Hemmerling J, Von Eiff C, Leverkus FW. Efficacy of PPV23 in preventing pneumococcal pneumonia in adults at increased risk - A systematic review and meta-analysis. Chalmers JD, ed. *PLoS One*. 2016;11(1):e0146338. doi:10.1371/journal.pone.0146338.

285. Falkenhorst G, Remschmidt C, Harder T, Hummers-Pradier E, Wichmann O, Bogdan C. Effectiveness of the 23-valent pneumococcal polysaccharide vaccine (ppv23) against pneumococcal disease in the elderly: Systematic review and meta-analysis. Ho PL, ed. *PLoS One*. 2017;12(1):e0169368. doi:10.1371/journal.pone.0169368.

286. Bonten MJM, Huijts SM, Bolkenbaas M, et al. Polysaccharide Conjugate Vaccine against Pneumococcal Pneumonia in Adults. *N Engl J Med*. 2015;372(12):1114-1125. doi:10.1056/NEJMoa1408544.

287. Suzuki M, Dhoubhadel BG, Ishifuji T, et al. Serotype-specific effectiveness of 23-valent pneumococcal polysaccharide vaccine against pneumococcal pneumonia in adults aged 65 years or older: a multicentre, prospective, test-negative design study. *Lancet Infect Dis*. 2017;17(3):313-321. doi:10.1016/S1473-3099(17)30049-X.

288. Streeter AJ, Lin NX, Crathorne L, et al. Adjusting for unmeasured confounding in non-randomised longitudinal studies: a methodological review. *J Clin Epidemiol*. 2017. doi:10.1016/j.jclinepi.2017.04.022.

289. Pletz MW, Kamradt T, Welte T. Time to follow up when comparing studies of pneumococcal vaccines. *Lancet Infect Dis*. 2017;17(3):244-246. doi:10.1016/S1473-3099(17)30051-8.

290. Millett ERC, Quint JK, Smeeth L, Daniel RM, Thomas SL. Incidence of community-acquired lower respiratory tract infections and pneumonia among older adults in the United Kingdom: a population-based study. Heimesaat MM, ed. *PLoS One*. 2013;8(9):e75131. doi:10.1371/journal.pone.0075131.

291. Skull SA, Andrews RM, Byrnes GB, et al. ICD-10 codes are a valid tool for identification of pneumonia in hospitalized patients aged > or = 65 years. *Epidemiol Infect*. 2008;136(2):232-240. doi:10.1017/S0950268807008564.

292. Clegg A, Bates C, Young J, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age Ageing*. 2016;45(3):353-360. doi:10.1093/ageing/afw039.

293. Rosenbaum PR, Rubin DB. Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *Am Stat*. 1985;39(1):33-38. http://amstat.tandfonline.com/doi/abs/10.1080/00031305.1985.10479383#.Uq17taUYLws. Accessed December 15, 2013.

294. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol*. 2006;163(12):1149-1156. doi:10.1093/aje/kwj149.

295. Quality Outcomes Framework. https://digital.nhs.uk/Quality-and-Outcomes-Framework/QOF.

296. García Rodríguez LA, Pérez Gutthann S. Use of the UK General Practice Research Database for pharmacoepidemiology. *Br J Clin Pharmacol*. 2002;45(5):419-425. doi:10.1046/j.1365-2125.1998.00701.x.

297. Kawakami K, Kishino H, Kanazu S, et al. Revaccination with 23-valent pneumococcal polysaccharide vaccine in the Japanese elderly is well tolerated and elicits immune responses. *Vaccine*. 2016;34(33):3875-3881. doi:10.1016/j.vaccine.2016.05.052.

298. Patterson S, Webber C, Patton M, et al. A post hoc assessment of duration of protection in CAPiTA (Community Acquired Pneumonia immunization Trial in Adults). *Trials Vaccinol*. 2016;5:92-96. doi:10.1016/j.trivac.2016.04.004.

299. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81(3):515-526. doi:10.1093/biomet/81.3.515.

300. World Health Organization. Vaccines against influenza WHO position paper –

November 2012. *Wkly Epidemiol Rec*. 2012;47(87):461-476. doi:10.1371/jour.

301. Demicheli V, Jefferson T, Di Pietrantonj C, et al. Vaccines for preventing influenza in the elderly. *Cochrane Database Syst Rev*. 2018;(2). doi:10.1002/14651858.CD004876.pub4.

302. Ludwig A, Lucero-Obusan C, Schirmer P, Winston C, Holodniy M. Acute cardiac injury events ≤30 days after laboratory-confirmed influenza virus infection among U.S. veterans, 2010-2012. *BMC Cardiovasc Disord*. 2015;15(109). doi:10.1186/s12872-015-0095-0.

303. Barnes M, Heywood AE, Mahimbo A, Rahman B, Newall AT, MaCintyre CR. Acute myocardial infarction and influenza: A meta-analysis of case-control studies. *Heart*. 2015;101(21):1738-1747. doi:10.1136/heartjnl-2015-307691.

304. Blackburn R, Zhao H, Pebody R, Hayward A, Warren-Gash C. Laboratory-Confirmed Respiratory Infections as Predictors of Hospital Admission for Myocardial Infarction and Stroke: Time-Series Analysis of English Data for 2004–2015. *Clin Infect Dis*. 2018;67(1):8-17. doi:10.1093/cid/cix1144.

305. Kwong JC, Schwartz KL, Campitelli MA, et al. Acute Myocardial Infarction after Laboratory-Confirmed Influenza Infection. *N Engl J Med*. 2018;378(4):345-353. doi:10.1056/NEJMoa1702090.

306. MacIntyre CR, Mahimbo A, Moa AM, Barnes M. Influenza vaccine as a coronary intervention for prevention of myocardial infarction. *Heart*. 2016;102(24):1953-1956. doi:10.1136/heartjnl-2016-309983.

307. Sager HB, Koenig W. Acute inflammation and long-term cardiovascular risk: Identifying an unrecognised vulnerable gap. *Eur J Prev Cardiol*. 2017;24(18):1956-1957. doi:10.1177/2047487317736869.

308. Clar C, Oseni Z, Flowers N, Keshtkar-Jahromi M, Rees K. Influenza vaccines for preventing cardiovascular disease. *Cochrane database Syst Rev*. 2015;5:CD005050. doi:10.1002/14651858.CD005050.pub3.

309. Gurfinkel EP, Leon De La Fuente R, Mendiz O, Mautner B. Flu vaccination in acute coronary syndromes and planned percutaneous coronary interventions (FLUVACS) Study: One-year follow-up. *Eur Heart J*. 2004;25(1):25-31.

doi:10.1016/j.ehj.2003.10.018.

310. Ciszewski A, Bilinska ZT, Brydak LB, et al. Influenza vaccination in secondary prevention from coronary ischaemic events in coronary artery disease: FLUCAD study. *Eur Heart J*. 2008;29(11):1350-1358. doi:10.1093/eurheartj/ehm581.

311. Chiang MH, Wu HH, Shih CJ, Chen YT, Kuo SC, Chen TL. Association between influenza vaccination and reduced risks of major adverse cardiovascular events in elderly patients. *Am Heart J*. 2017;193:1-7. doi:10.1016/j.ahj.2017.07.020.

312. Udell JA, Zawi R, Bhatt DL, et al. Association between influenza vaccination and cardiovascular outcomes in high-risk patients: a meta-analysis. *JAMA*. 2013;310(16):1711-1720. doi:10.1001/jama.2013.279206.

313. Mann AG, Mangtani P, Russell CA, Whittaker JC. The impact of targeting all elderly persons in England and Wales for yearly influenza vaccination: Excess mortality due to pneumonia or influenza and time trend study. *BMJ Open*. 2013;3(8):e002743. doi:10.1136/bmjopen-2013-002743.

314. Hsu S-Y, Chen F-L, Liaw Y-P, Huang J-Y, Nfor ON, Chao D-Y. A Matched Influenza Vaccine Strain Was Effective in Reducing the Risk of Acute Myocardial Infarction in Elderly Persons: A Population-Based Study. *Medicine (Baltimore)*. 2016;95(10):e2869. doi:10.1097/MD.0000000000002869.

315. Boni MF, Gog JR, Andreasen V, Christiansen FB. Influenza drift and epidemic size: the race between generating and escaping immunity. *Theor Popul Biol*. 2004;65(2):179-191. http://cat.inist.fr/?aModele=afficheN&cpsidt=15654780. Accessed December 5, 2012.

316. Veljkovic V, Glisic S, Veljkovic N, et al. Influenza vaccine as prevention for cardiovascular diseases: Possible molecular mechanism. *Vaccine*. 2014;32(48):6569-6575. doi:10.1016/j.vaccine.2014.07.007.

317. Prentice RL, Williams BJ, Peterson A V. On the regression analysis of multivariate failure time data. *Biometrika*. 1981;68(2):373-379. doi:10.1093/biomet/68.2.373.

318. Jackson LA, Jackson ML, Nelson JC, Neuzil KM, Weiss NS. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. *Int J Epidemiol*. 2006;35(2):337-344.

319. Nelson JC, Jackson ML, Weiss NS, Jackson LA. New strategies are needed to improve the accuracy of influenza vaccine effectiveness estimates among seniors. *J Clin Epidemiol*. 2009;62(7):687-694.

320. Groenwold RHH, Hoes AW, Hak E. Impact of influenza vaccination on mortality risk among the elderly. *Eur Respir J Off J Eur Soc Clin Respir Physiol*. 2009;34(1):56-62. doi:10.1183/09031936.00190008.

321. Kwong JC, Campitelli MA, Gubbay JB, et al. Vaccine effectiveness against laboratory-confirmed influenza hospitalizations among elderly adults during the 2010-2011 season. *Clin Infect Dis*. 2013;57(6):820-827. doi:10.1093/cid/cit404.

322. Streeter AJ, Lin NX, Crathorne L, et al. Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review. *J Clin Epidemiol*. 2017;87:23-34. doi:10.1016/j.jclinepi.2017.04.022.

323. Streeter AJ, Masoli JA, Blé A, Melzer D, Henley WE. Pneumococcal Vaccine Effectiveness and Its Interaction With Age: A UK Population Based Study in Older Adults [400]. *Pharmacoepidemiol Drug Saf*. 2017;26(52):3-636. doi:10.1002/pds.4275.

324. Streeter AJ, Henley WE. [22] Evidence from a quasi-experimental study for the effectiveness of the influenza vaccination against myocardial infarction in UK adults aged at least 65y. In: *Pharmacoepidemiology and Drug Safety*. Vol 27. Wiley-Blackwell; 2018:3-521. doi:10.1002/pds.4629.

325. Henley WE, Streeter AJ. [23] Real-world effectiveness of influenza vaccination in older adults in the UK from 1997-2012: A quasi-experimental cohort study. In: *Pharmacoepidemiology and Drug Safety*. Vol 27. Wiley-Blackwell; 2018:3-521. doi:10.1002/pds.4629.

326. Mangtani P, Cumberland P, Hodgson CR, Roberts JA, Cutts FT, Hall AJ. A cohort study of the effectiveness of influenza vaccine in older people, performed using the United Kingdom general practice research database. *J*

*Infect Dis*. 2004;190(1):1-10. doi:10.1086/421274.

327. Darvishian M, Bijlsma MJ, Hak E, van den Heuvel ER. Effectiveness of seasonal influenza vaccine in community-dwelling elderly people: a meta-analysis of test-negative design case-control studies. *Lancet Infect Dis*. 2014;14(12):1228-1239. doi:10.1016/S1473-3099(14)70960-0.

328. Feise R. Do multiple outcome measures require p-value adjustment? *BMC Med Res Methodol*. 2002;2(1):8. doi:10.1186/1471-2288-2-8.

329. Warren-Gash C, Smeeth L, Hayward AC. Influenza as a trigger for acute myocardial infarction or death from cardiovascular disease: a systematic review. *Lancet Infect Dis*. 2009;9(10):601-610. doi:10.1016/S1473-3099(09)70233-6.

330. Ruane L, Buckley T, Hoo SYS, et al. Triggering of acute myocardial infarction by respiratory infection. *Intern Med J*. 2017;47(5):522-529. doi:10.1111/imj.13377.

331. Gwini SM, Coupland CAC, Siriwardena AN. The effect of influenza vaccination on risk of acute myocardial infarction: Self-controlled case-series study. *Vaccine*. 2011;29(6):1145-1149. doi:10.1016/j.vaccine.2010.12.017.

332. Sribhutorn A, Phrommintikul A, Wongcharoen W, Chaikledkaew U, Eakanunkul S, Sukonthasarn A. The Modification Effect of Influenza Vaccine on Prognostic Indicators for Cardiovascular Events after Acute Coronary Syndrome: Observations from an Influenza Vaccination Trial. *Cardiol Res Pract*. 2016;2016:1-9. doi:10.1155/2016/4097471.

333. Mohseni H, Kiran A, Khorshidi R, Rahimi K. Influenza vaccination and risk of hospitalization in patients with heart failure: A self-controlled case series study. *Eur Heart J*. 2017;38(5):326-333. doi:10.1093/eurheartj/ehw411.

334. Osterhaus A, Abdullah Brooks W, Broberg E, Raina MacIntyre C, Capua I. Why should influenza be a public health priority? In: *Vaccine*. Vol 33. Elsevier; 2015:7022-7025. doi:10.1016/j.vaccine.2015.08.049.

335. Vardeny O, Solomon SD. Influenza vaccination: a one-shot deal to reduce cardiovascular events. *Eur Heart J*. 2016;38(5):ehw560.

doi:10.1093/eurheartj/ehw560.

336. Fröbert O, Götberg M, Angerås O, et al. Design and rationale for the Influenza vaccination After Myocardial Infarction (IAMI) trial. A registry-based randomized clinical trial. *Am Heart J*. 2017;189:94-102. doi:10.1016/j.ahj.2017.04.003.

337. Warren-Gash C, Udell JA. Respiratory Tract Infections, Nonsteroidal Anti-inflammatory Drugs and Acute Myocardial Infarction: Is Understanding Interaction Between Risk Factors the Key to Personalizing Prevention? *J Infect Dis*. 2017;215(4):497-499. doi:10.1093/infdis/jiw604.

338. Fedson DS. Pandemic Influenza: A Potential Role for Statins in Treatment and Prophylaxis. *Clin Infect Dis*. 2006;43(2):199-205. doi:10.1086/505116.

339. Black S, Nicolay U, Del Giudice G, Rappuoli R. Influence of Statins on Influenza Vaccine Response in Elderly Individuals. *J Infect Dis*. 2016;213(8):1224-1228. doi:10.1093/infdis/jiv456.

340. McLean HQ, Chow BDW, VanWormer JJ, King JP, Belongia EA. The Effect of Statin Use on Influenza Vaccine Effectiveness. *J Infect Dis*. 2016;214(8):1150-1158. doi:10.1093/infdis/jiw335.

341. Omer SB, Phadke VK, Bednarczyk RA, Chamberlain AT, Brosseau JL, Orenstein WA. Impact of Statins on Influenza Vaccine Effectiveness Against Medically Attended Acute Respiratory Illness. *J Infect Dis*. 2016;213(8):1216-1223. doi:10.1093/infdis/jiv457.

342. Izurieta HS, Chillarige Y, Kelman JA, et al. Statin use and risks of influenza-related outcomes among older adults receiving standard-dose or high-dose influenza vaccines through Medicare during 2010-2015. *Clin Infect Dis*. February 2018. doi:10.1093/cid/ciy100.

343. *The REporting of Studies Conducted Using Observational Routinely-Collected Health Data (RECORD) Statement for Pharmacoepidemiology (RECORD-PE)*.; 2018. http://www.record-statement.org. Accessed July 14, 2018.

344. Macfarlane J. Community-acquired pneumonia. *Br J Dis Chest*. 1987;81:116-127. doi:10.1016/0007-0971(87)90128-8.

345. Ashton C, Bajekal M, Raine R. Quantifying the contribution of leading causes of death to mortality decline among older people in England, 1991-2005. *Health Stat Q*. 2010;(45):100-127. doi:10.1057/hsq.2010.6.

346. Lim WS, Baudouin S V, George RC, et al. BTS guidelines for the management of community acquired pneumonia in adults: update 2009. *Thorax*. 2009;64 Suppl 3(Suppl_3):iii1-55. doi:10.1136/thx.2009.121434.

347. Thompson WW, Shay DK, Weintraub E, et al. Influenza-associated hospitalizations in the United States. *JAMA*. 2004;292(11):1333-1340. doi:10.1001/jama.292.11.1333.

348. Thompson WW. Mortality Associated With Influenza and Respiratory Syncytial Virus in the United States. *JAMA*. 2003;289(2):179. doi:10.1001/jama.289.2.179.

349. Jefferson T, Di Pietrantonj C, Al-Ansary LA, Ferroni E, Thorning S, Thomas RE. Vaccines for preventing influenza in the elderly. *Cochrane database Syst Rev*. 2010;(2):CD004876. doi:10.1002/14651858.CD004876.pub3.

350. Govaert TM, Thijs CT, Masurel N, Sprenger MJ, Dinant GJ, Knottnerus JA. The efficacy of influenza vaccination in elderly individuals. A randomized double-blind placebo-controlled trial. *JAMA*. 1994;272(21):1661-1665. http://www.ncbi.nlm.nih.gov/pubmed/7966893. Accessed February 11, 2013.

351. Lamontagne F, Garant M-P, Carvalho J-C, Lanthier L, Smieja M, Pilon D. Pneumococcal vaccination and risk of myocardial infarction. *CMAJ*. 2008;179(8):773-777. doi:10.1503/cmaj.070221.

352. Tseng HF, Slezak JM, Quinn VP, Sy LS, Van den Eeden SK, Jacobsen SJ. Pneumococcal vaccination and risk of acute myocardial infarction and stroke in men. *JAMA*. 2010;303(17):1699-1706. doi:10.1001/jama.2010.529.

353. Siriwardena AN, Gwini SM, Coupland CAC. Influenza vaccination, pneumococcal vaccination and risk of acute myocardial infarction: matched case-control study. *CMAJ*. 2010;182(15):1617-1623. doi:10.1503/cmaj.091891.

354. Rubin DB, Thomas N. Matching Using Estimated Propensity Scores: Relating

Theory to Practice. *Biometrics*. 1996;52(1):249-264.

355. D'Agostino RB. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med*. 1998;17(19):2265-2281. http://www.ncbi.nlm.nih.gov/pubmed/9802183. Accessed October 10, 2013.

356. Nevalainen J, Kenward MG, Virtanen SM. Missing values in longitudinal dietary data: a multiple imputation approach based on a fully conditional specification. *Stat Med*. 2009;28(29):3657-3669. doi:10.1002/sim.3731.

357. White IR, Royston P. Imputing missing covariate values for the Cox model. *Stat Med*. 2009;28(15):1982-1998. doi:10.1002/sim.3618.