Quantifying the risk of Hodgkin lymphoma in symptomatic primary care patients over 40: a case-control study using electronic records.

Elizabeth A Shephard, PhD, CPsychol, Research Fellow

Richard D Neal, PhD FRCGP, Professor of Primary Care Medicine

Peter Rose, MD FRCGP, Senior Clinical Researcher

Fiona M Walter, MD, FRCGP, GP & Clinician Scientist

William T Hamilton, MD, FRCP, FRCGP, Professor of Primary Care Diagnostics

1 University of Exeter Medical School

College House

St Luke’s Campus

Magdalen Road

Exeter

EX1 2LU

2 North Wales Centre for Primary Care Research

Bangor University

Gwenfro Unit 5

Wrexham Technology Park

Wrexham

LL13 7YP

3 Department of Primary Care Health Sciences

University of Oxford

New Radcliffe House
2nd Floor
Radcliffe Observatory Quarter
Woodstock Road
Oxford
OX2 6GG

"Department of Public Health & Primary Care
University of Cambridge
Strangeways Research Laboratory
Cambridge
CB1 8SR

Correspondence to Dr Shephard

Email: E.A.Shephard@exeter.ac.uk
ABSTRACT

Background: In the UK, approximately five people are diagnosed with Hodgkin lymphoma (HL) daily. One tenth of diagnoses are in the over 75s.

Aim: To establish a symptom profile of HL and quantify their risk in primary care patients over 40.

Design and setting: Matched case-control study using Clinical Practice Research Datalink patient records.

Methods: Putative clinical features of HL were identified in the year before diagnosis. Results were analysed using conditional logistic regression and positive predictive values (PPVs) calculated for the consulting population.

Results: 283 patients aged ≥40 years, diagnosed with HL between 2000 and 2009, and 1,237 age, sex and general practice-matched controls were studied. Six features were independently associated with HL: lymphadenopathy, odds ratio 2.80 (95% confidence interval 25,3100), head and neck mass not described as lymphadenopathy 2.60 (21,3200), other mass 12 (4.4,35), thrombocytosis 6.0 (2.6,14), raised inflammatory markers 5.2 (3.0,9.0), and low full blood count 2.8 (1.6,4.8). Lymphadenopathy per se has a positive predictive value (PPV) of 5.6% for HL in patients over 60 years.

Conclusions: Consistent with secondary care findings, lymphadenopathy is the clinical feature with the highest risk of HL in primary care and warrants urgent investigation.

Keywords: Hodgkin lymphoma; Primary Health Care; diagnosis
HOW THIS FITS IN

What is already known on this subject?

- Over 40% of HL patients visit their GP three or more times before being referred to a specialist.

- Lymphadenopathy is a known feature of HL and a primary indicator for referral in the NICE guidance for haematological cancer though based on little primary care research. In this study...

What this study adds

- Patients over the age of sixty with lymphadenopathy have a 5.6% risk of having HL.

- The current guidance advising a six weeks wait for unexplained lymphadenopathy may be unnecessarily long.
INTRODUCTION

Lymphoma is a cancer of the lymphatic system. Approximately 10% of lymphomas are Hodgkin lymphoma (HL), the remaining being Non-Hodgkin lymphoma (NHL). Annually in the UK, more than 1,800 people are diagnosed with HL with 300 deaths. (1) It has a bimodal pattern of incidence by age, with a peak at ages 20-24 and a second peak in women aged 70-74 and men aged 75-79. (1) More than 80% of HL cases survive for at least 5 years.(2)

The UK compares unfavourably to the European average for cancer survival. Between 1995 and 1999, eleven percent of HL deaths were deemed avoidable. (3) The main source of delay in HL occurs during diagnosis: over 40% of HL patients consulted their GP with symptoms three or more times before being referred.(4)

Current guidance for suspected haematological malignancies lists weight loss, abdominal pain, fever, fatigue and lymphadenopathy as warranting further investigation. A full blood count, blood film and inflammatory markers are recommended for people with fatigue or unexplained lymphadenopathy. Immediate investigation is recommended for persistent lymphadenopathy for over 6 weeks, lymph nodes of over 2cm diameter, increasing size or widespread nature, or accompanied by weight loss, splenomegaly or night sweats.(5)

Nearly 87% of HL patients have reported symptoms before their diagnosis, though these have rarely been studied in primary care, the setting where most patients initially present. (6) One primary care study investigated fifteen years period before a lymphoma diagnosis and reported increased infections in HL patients.(7) In secondary care studies, painless lymphadenopathy has long been recognised as a feature of lymphoma, with painful lymphadenopathy also recently noted.(6) Abdominal pain, fatigue, stomach/bowel problems, infection, back pain and pain on drinking alcohol are also reported, though the alcohol-induced pain findings reported almost 50 years ago have not
been reliably re-reported. (6, 8, 9) General symptoms, sometimes called B symptoms, of intermittent fever, weight loss, and night sweats are also recognised. (10) Associations with anaemia up to two years before diagnosis have been reported (odds ratio 1.26 (Confidence interval 1.00,1.59) per standard deviation reduction from the mean haemoglobin). (11)

This study aimed to identify and quantify the early clinical features (symptoms, signs and abnormal investigations) of HL in primary care, with the aim of expediting the selection of patients for definitive investigation.
METHODS

This study is part of a project looking at the clinical profile of 13 common cancers among primary care patients aged forty years and over (see companion paper in this issue). The matched case-control design used UK primary care electronic patient records from the Clinical Practice Research Datalink (CPRD). The CPRD contains anonymised patient data collected from over 680 general practices, covering 8.8% of the UK population. Data includes primary care clinical events such as symptom reporting, investigations, diagnoses, prescriptions and specialist referrals. The CPRD applies stringent quality standards for data entry.

Cases and controls

A list of 172 lymphoma codes (available from authors) was collated; twenty-nine codes applied to HL. Cases were selected if aged ≥40 years and diagnosed with HL between January 2000 and December 2009. Up to five age, sex and practice controls were matched to each case. The first HL code was taken as the date of diagnosis. The index date for controls matched their case’s diagnosis date. Exclusion criteria were: cases with NHL, mycosis fungoides or Sézary syndrome and their matched controls; any case or control with less than one year of records before the index date; cases without controls; controls with HL; and controls who had not sought medical care after registration.

Selection of putative clinical variables

A list of potential HL clinical features (symptoms, signs and abnormal investigations) was compiled from existing literature and online patient support group threads. Google Scholar, PubMed, and EBSCO were used with the search terms ‘Hodgkin lymphoma symptoms’, ‘Hodgkin lymphoma primary care’, and ‘early signs/indications/symptoms of Hodgkin lymphoma’. Over 100,000 medical codes are recorded in the CPRD; several relevant codes relate to each feature, which we collated into symptom libraries. Instances were identified in the year before the index date, and variables found in at least 2% of cases were retained. Recording bias was tested on a condition thought to
have no association with HL – varicose veins. Abnormal investigation results were defined as the patient having a test value falling outside their local laboratory’s normal range. Patients with a normal laboratory result were grouped with those without testing.

**Composite variables**

Some tests were grouped together. (12, 13) The raised inflammatory markers variable comprised any of: abnormal erythrocyte sedimentation rate, plasma viscosity, or C-reactive protein. Abnormal liver function tests reflected a raised value of any of the hepatic enzymes. Low full blood count included any of: low haemoglobin, low white cell count or low platelets. Three categories of mass variables were compiled. First we identified masses in the head or neck incorporating cervical lymphadenopathy; next we identified masses labelled as lymphadenopathy (this included generalised lymphadenopathy, and lymphadenopathy with no site mentioned); finally we included all masses from elsewhere in the body as ‘other mass’. Exact information relating to the size, location or duration of mass was unavailable. To indirectly estimate the duration of a patient reporting a mass variable (duration is rarely recorded, so direct measurement was impossible) the first and last report of any of the three mass variables in the year before diagnosis was identified. Those masses having an apparent duration of at least 42 days were reported in a sub-analysis (42 days is the duration of lymphadenopathy recommended for investigation in current NICE guidance).(5)

**Analysis and statistical methods**

The main form of analysis was conditional logistic regression. (12, 13) Non-parametric methods were used as the data was not normally distributed. Univariable analysis was performed on all features present in at least 2% of either cases or controls. Those features associate with HL at a p-value ≤0.1 were then grouped into small clinically coherent groups (such as back pain, abdominal pain and chest pain) for multivariable analyses. Those variables with an association at a p-value of ≤0.05
entered the final stage of modelling, with a p-value threshold of 0.01 used for retention in the final model. Variables omitted at these earlier stages were checked against the final model, and restored if a likelihood ratio test comparing the models had a p-value ≤0.01. Clinically plausible interaction terms were added to the final model and retained if their p-value was also ≤0.01.

Risk estimates in the form of positive predictive values (PPVs) were calculated using Bayes’ theorem (prior odds x likelihood ratio = posterior odds). Prior odds were calculated from the age-specific national incidence of HL for 2008, expressed as odds. PPVs for pairs of features and repeated attendances for the same feature were calculated where indicated. PPVs were estimated for consulting patients only; thus, the posterior odds were divided by 0.885 as 161 (11.5%) of 1,398 eligible controls were non-consulters (see Figure 2).

No sub-analysis by histological subtype was possible due to the small sample size.

Power calculation
Due to the fixed number of participant numbers received from the CPRD (300 cases and 1350 controls), power calculations were used instead of sample size calculations. The numbers provided >86% power (5% two-sided alpha) to detect a difference in a rare variable present in 4% of cases compared to 1% of controls. For a commoner variable, the study had >83% power to detect a change in prevalence of 20% in cases and 13% in controls. Data analysis was conducted using Stata software, version 13.1.
RESULTS

The CPRD provided 28,502 patients (4,799 cases; 23,703 controls). This included all sub-types of lymphoma cases and their controls. Application of the exclusion criteria is shown in Figure 1, leading to a final number of 1,520 in this study (283 HL cases; 1,237 controls).

Patient demographic and consultation information is given in Table 1. Cases consulted significantly more frequently than controls in the year before diagnosis (p=<0.001; ranksum test).

Clinical features

Forty-three symptoms and twenty-two abnormal test results were considered initially. Six remained significant in the final model: lymphadenopathy, head and neck mass not described as lymphadenopathy, other mass, thrombocytosis, raised inflammatory markers, and low full blood count. Their frequencies, univariable likelihood ratios and multivariable odds ratios are shown in Table 2. Of the three B symptoms reported with lymphoma- fever, sweating and weight loss- both sweating and weight loss were frequent enough to proceed to multivariable analysis. There were only 3 (1.1%) cases and 6 (0.5%) controls with recorded fever. The proportion of patients with varicose veins did not differ between cases and controls (p<0.62) – suggesting any recording bias was minimal. In total, 9 (2.1%) cases and 0 controls had multiple consultations with one of the mass variables at least 42 days apart. No interaction terms, including with gender, were found. From the 283 cases, 187 (66%) had at least one of the final model features from Table 2 recorded.
Positive predictive values

PPVs were not calculated if fewer than 5 cases had the feature. We present here the PPV figures for the over 60s – for two main reasons. First, they were the majority of the cohort (mirroring the expected epidemiology); second, we have published an over 60s Risk Assessment Tool for NHL (see companion paper in this issue). By estimating figures for the same age group, the PPV values can be added together, to give estimated risks for each symptom. Where fewer than 10 cases or controls had the combined features, CIs were omitted. In the over 60s age group, lymphadenopathy, head and neck mass and other mass as single symptoms had PPVs of 5.6%, 2.3% and 0.03%, respectively. Lymphadenopathy combined with a low full blood count or raised inflammatory markers, produced PPVs of 2.5% and 2.2% respectively. Thrombocytosis as a single feature produced a small PPV of 0.04% (CI 0.02,0.07). The remaining PPVs were either under 0.05% or could not be calculated due to small numbers.
DISCUSSION

Summary

This is the first study to identify and quantify the clinical features of primary care HL in patients aged 40 years and over. Three symptoms and three abnormal investigations were associated with HL in the whole cohort. The highest risk estimate for HL in the over 60s was lymphadenopathy, though masses elsewhere in the body had lower risks. Using the whole cohort, risk estimates were between 0.9% and 2.8% when combining raised inflammatory markers or low full blood count with the three mass variables.

Strengths and limitations

This study is the first to investigate the primary care features of HL patients, separately from NHL, in the year before their cancer diagnosis. By doing so, we could have identified important differences in symptom profiles (though no surprising differences were identified). Equally, it is possible to sum the PPVs within the same age band, to give an overall ‘lymphoma-risk’ for specific features. The advantage of using CPRD data is well known. (15, 16) The data is generalizable and representative of the UK population. Our comprehensive methods for identifying putative features of HL, including using online support forums to search for patient-reported symptoms, mean it is unlikely that relevant features were omitted. Furthermore, the study’s primary care setting is important. The clinical problem of whom to select for cancer investigation resides in primary care, so requires primary care research.

This study is part of a programme looking at the cancer prodrome in primary care patients forty years of age and over. The starting age of forty was chosen for two reasons. Firstly, for most cancers, the median age of cancer diagnosis is in the young seventies. Additionally, cancer in the under 40s disproportionally represent familial syndromes, with the possibility of atypical presentations. Thus the programme as a whole selected a lower age of 40 – a figure that has been
problematic only for Hodgkin’s lymphoma and testis. Despite this, there is no published evidence to suggest that the clinical features of HL differ by age, though this remains a possibility. Survival from HL is superior in the young, though this does not equate to their being a different symptom profile.

(17) Our sample was relatively small, thus precluding any sub-analysis and restricting the number of risk estimates which we were able to reliably generate. Data recording may not be consistent amongst GPs, though this is somewhat mitigated through the limited number of codes available to the GPs, who generally selected a particular ‘dominant’ code for each symptom. Additionally, relevant clinical information can be recorded in a hidden ‘free text’ section, opening the possibility of preferential recording for cases over controls.(18) Cases have more chances to report symptoms due to their higher consultation pattern; however, our test for recording bias did not support this. Finally, as in our previous studies, we overcame the problem of estimating PPVs from a case-control study design by calculating the prior odds of HL from registry data.(12, 19)

Comparison with existing literature
Consistent with secondary care literature, our main finding was lymphadenopathy as conferring the highest risk for HL; head and neck masses had a lesser but considerable association.(6, 20, 21) These results are concordant with those found in our companion paper on NHL. Cervical lymphadenopathy in primary care has been investigated in three old studies.(22-24) In the first, no malignancies were found from a group of 80 primary care patients with lymphadenopathy, 44% of whom had isolated cervical node enlargement.(24) A second study of 249 patients also found no malignancies, and in the majority of patients no final diagnosis was established.(22) These two studies had younger patients; in such patients, lymphadenopathy of infectious origin will predominate. Our large PPVs were seen only in the over 60s, when infection becomes relatively less important. The third study examined GP referrals for lymph node biopsy. 29 malignancies were found in 82 referred patients, with a prior probability for lymphadenopathy presenting to primary care of 1.1% calculated from these results.(23) All three studies emphasized the ability of GPs to identify lymphadenopathy of
malignant origin. The association with anaemia has been reported before in the secondary care literature. (25)

We found no association with the so-called B symptoms – this may simply reflect our relatively small sample size, or the exclusion of the younger patients. This is part of a bigger problem in cancer diagnostic research: the symptoms that span several possible cancer sites, such as weight loss or thrombocytosis. At a single cancer site level, the risk may be small (even so small it cannot be identified) but across cancer as a whole it may be important. To study these features, cohort studies starting with the symptom will have to be used; this has been done for jaundice and hypercalcaemia, and is under way for thrombocytosis.(26, 27)

**Implications for practice**

The main finding in this study is the importance of lymphadenopathy and head and neck masses. When combined with NHL, the risks with these symptoms are high, and warrant serious consideration of lymphoma. We were also unable to examine persistence of the mass directly, as duration of symptoms is poorly recorded in medical records including the CPRD.(28) Our proxy for duration, two mass codes 6 weeks apart, had a PPV of 1.3% (0.0,94), in the over 60s. The word ‘unexplained’ is used in referral guidance, including the latest revision.(5) As with the NHL paper, it is likely that many of the masses were unexplained, as GPs prefer to document diagnoses where possible. Thus, much of the ‘explained’ lymphadenopathy may well be unrecorded.

When the results of this study are added to those for NHL in the companion paper (ref please subeditor), the PPVs rise to 18.6% for either NHL or HL for lymphadenopathy, 4.6% for head and neck mass, and 1.1% for other mass elsewhere. Thus the risk estimates for the three mass variables indicate that the default decision should be referral of patients over 60 with lymphadenopathy or a head and neck mass, unless there is a clear alternative explanation.
Conclusion

Lymphadenopathy and head and neck masses in adults over the age of 60 are strong predictors of both HL and NHL and warrant urgent investigation, particularly if they have been present for six weeks or more. No blood test or other symptoms change that. This largely accords with current guidance, though it could be argued that the need to wait six weeks – to allow resolution or an alternative diagnosis to emerge – is unnecessarily long.
**Funding:** This article presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research Programme (Grant Reference Number RP-PG-0608-10045). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. FMW is funded by a NIHR Clinician Scientist award. RDN is part funded by Public Health Wales and Betsi Cadwaladr University Health Board.

**Ethical approval:** Independent Scientific Advisory Committee – protocol 09-110

**Competing interests:** WH is clinical lead on the ongoing revision of the NICE guidance on investigation of suspected cancer. His contribution to this article is in a personal capacity, and is not to be interpreted as representing the view of the Guideline Development Group, or of NICE itself. PR reports personal fees from GP Update Ltd, outside the submitted work. Other than this, no competing interests.

**Acknowledgments:** The authors would like to acknowledge the contribution to the research presented in this paper made by the Discovery Programme Steering Committee comprising: Roger Jones (chair); Jonathan Banks; Alison Clutterbuck; Jon Emery; Joanne Hartland; Sandra Hollinghurst; Maire Justice; Jenny Knowles; Helen Morris; Tim Peters; Greg Rubin.
REFERENCES


3. Abdel-Rahman M, Stockton D, Rachet B, Hakulinen T, Coleman MP. What if cancer survival in Britain were the same as in Europe: how many deaths are avoidable? Br J Cancer. 2009;101(S2):S115-S24.


Table 1. Patient demographics and consultation rates in the year before diagnosis

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male (n=158)</td>
<td>Female (n=125)</td>
</tr>
<tr>
<td>Median (IQR) age in years at diagnosis</td>
<td>61 (53-71)</td>
<td>65 (57-73)</td>
</tr>
<tr>
<td>Median (IQR) number of consultations</td>
<td>14 (8-23)</td>
<td>17 (11-24)</td>
</tr>
</tbody>
</table>
Table 2. Features of HL (all ages)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Cases, n (%)</th>
<th>Controls, n (%)</th>
<th>Likelihood ratio (95% CI)</th>
<th>Odds ratio in multivariable analysis (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=283</td>
<td>n=1237</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Symptoms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lymphadenopathy</td>
<td>51 (18)</td>
<td>0 (0)</td>
<td>223* (31-1606)</td>
<td>282* (25 to 3123)</td>
</tr>
<tr>
<td>Head and neck lump</td>
<td>30 (11)</td>
<td>0 (0)</td>
<td>131* (18-958)</td>
<td>261* (21 to 3171)</td>
</tr>
<tr>
<td>Lump</td>
<td>19 (7)</td>
<td>15 (1)</td>
<td>5.5 (2.9-11)</td>
<td>12 (4.4 to 35)</td>
</tr>
<tr>
<td><strong>Investigations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raised inflammatory markers</td>
<td>105 (37)</td>
<td>66 (5)</td>
<td>7.0 (5.3-9.2)</td>
<td>5.2 (3.0 to 9.0)</td>
</tr>
<tr>
<td>Low full blood count</td>
<td>98 (35)</td>
<td>88 (7)</td>
<td>4.9 (3.8-6.3)</td>
<td>2.8 (1.6 to 4.8)</td>
</tr>
<tr>
<td>Thrombocytosis</td>
<td>54 (19)</td>
<td>17 (1)</td>
<td>14 (8.2-24)</td>
<td>6.0 (2.6 to 14)</td>
</tr>
</tbody>
</table>

* For analysis purposes 1 control was added to lymphadenopathy and head and neck lump
**Figure 1: HLeclusion data**

**Total number**

*n=28,502*

**Controls**

*n=23,703*

- Control with only 1 line of data and no event date information
  *n=26*

- Excluded control of case with mycosis fungoides or Sézary syndrome
  *n=719*

- Excluded control of case with Non-Hodgkin lymphoma
  *n=21,556*

- HL only controls
  *n=1,402*

- Excluded control with Hodgkin lymphoma before 2000
  *n=3*

- Excluded control with Hodgkin lymphoma after 2000
  *n=1*

**Cases**

*n=4,799*

- Excluded case with mycosis fungoides or Sézary syndrome
  *n=147*

- Excluded case with Non-Hodgkin lymphoma
  *n=4,369*

- HL cases eligible for inclusion
  *n=283*

- Excluded case no controls
  *n=0*

**Controls eligible for inclusion**

*n=1,398*

- Excluded control no data in year pre index date
  *n=161*

**Total controls included**

*n=1,237*