
Article:
Crouch, Simon orcid.org/0000-0002-3026-2859, Simpson, Jill, Ansell, Pat et al. (6 more authors) (2011) Illness patterns prior to diagnosis of lymphoma : Analysis of UK medical records. Cancer Epidemiology. pp. 145-150. ISSN 1877-7821

https://doi.org/10.1016/j.canep.2010.08.003

Cancer Epidemiology xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

Illness patterns prior to diagnosis of lymphoma: Analysis of UK medical records

Simon Crouch [a,*], Jill Simpson [a], Pat Ansell [a], Eleanor Kane [a], Debra Howell [a], Alex Smith [a], Rob Newton [a], Andrew Jack [b], Eve Roman [a]

[a] Epidemiology and Genetics Unit, Department of Health Sciences, University of York, Area 3 Seebohm Rowntree Building, Heslington, York, YO10 5DD, UK[1]
[b] Haematological Malignancy Diagnostic Service, Leeds General Infirmary, Leeds, UK[2]

ARTICLE INFO

Article history:
Accepted 6 August 2010

Keywords:
Lymphoma
Infections
Epidemiology

ABSTRACT

Background: Increased understanding of the relationship between lymphomas and co-morbidities is likely to provide valuable insights into the natural history of these disorders. *Methods*: 761 Cases with lymphoma (310 diffuse-large B-cell [DLBCL]; 226 follicular [FL]; and 225 Hodgkin [HL]) and 761 unaffected age and sex matched controls were recruited and their histories of infection and non-infection diagnoses in primary care records were compared using negative binomial regression. *Results*: No differences were observed between the infectious illness patterns of DLBCL and FL cases and their matched controls over the 15 years preceding lymphoma diagnosis. A marked excess of infectious illness episodes was recorded for HL cases compared to their controls; evident at least a decade prior to HL diagnosis. For non-infectious consultations an excess of case over control visits emerged 4–6 years before DLBCL and FL diagnosis; no specific co-morbidity associations were found. No case–control differences for non-infectious conditions were apparent for HL. *Conclusion*: There are substantial variations in patterns of illness prior to diagnosis of the three lymphoma subtypes examined. The excess of infectious diagnoses prior to HL may point to underlying immune abnormality, but there was no suggestion of this for DLBCL and FL where a generalized excess of non-infectious conditions was evident.

1. Introduction

There is broad consensus that increased understanding about the nature of the relationship between lymphomas and other co-morbidities, particularly auto-immune and infectious conditions, is likely to provide valuable insights into the natural history of these lymphoproliferative disorders [1]. Immunosuppression, whether related to HIV infection or drug treatment, such as that experienced by renal transplant recipients, appears to be associated with a modest increase in risk of Hodgkin lymphoma (HL) and a greater increase in risk of certain types of non-Hodgkin lymphoma (NHL) [2,3]. Whilst a few subtypes of lymphoma are thought to be related to specific infections there is little evidence that this is true for the majority, but there is some support for the notion that non-specific infectious episodes several years prior to lymphoma diagnosis may signal disease initiation and/or progression [4].

In order to investigate the potential association between infectious and other immunological factors and subsequent lymphoma risk we systematically abstracted primary-health care medical records of patients enrolled in a UK lymphoma case–control study. We report here on the role of clinically diagnosed medical conditions (as recorded in primary care medical records) in the two commonest subtypes of NHL (diffuse-large B-cell lymphoma and follicular lymphoma) and HL.

2. Methods

Details of the UK population-based case–control study are described elsewhere in detail [4,5]. Briefly, cases comprised patients newly diagnosed with lymphoma (non-HIV-related) residing in pre-defined geographic areas and newly diagnosed with lymphoma before 65 years of age during 1998–2003. Diagnoses were confirmed pathologically and coded according to the World Health Organisation Classification [6]. For each case, one age and sex matched control was randomly selected from population registers. The overall response rate was 75% in cases and 71% in controls, which compares favourably with similar studies conducted elsewhere in the world [7].

The ability to access data from an individual's primary care records over their lifetime is a major feature of the UK National Health Service (NHS). For this reason, at interview subjects were asked to consent to access to their primary care records; and all of the information contained therein for the 15 years prior to diagnosis in cases (or pseudo-diagnosis in controls) was subsequently abstracted onto specially designed forms by trained research staff. For each contact with primary care, the information recorded included all illnesses diagnosed at each consultation by

* Corresponding author. Fax: +44 01904 321 899.
E-mail address: simon.crouch@egu.york.ac.uk (S. Crouch).
[1] www.egu.york.ac.uk.
[2] www.hmds.org.uk.

1877-7821/$ – see front matter © 2010 Published by Elsevier Ltd.
doi:10.1016/j.canep.2010.08.003

the patient's general practitioner (GP, i.e. their primary care physician), as well as all signs and symptoms with which they presented at the time, as well as resultant referrals to hospital or other specialist organizations, results of the investigations, and details of medicines or other prescribed therapies. All such contemporaneously recorded data were abstracted.

Data abstraction and data entry were structured around dated 'events'. Disease and drug coding was done centrally by experienced primary care research nurses, using a specially designed computerised system embedded within the data entry programme. Illnesses and symptoms were coded according to the International Statistical Classification of Diseases and Related Health Problems tenth revision (ICD-10) [8], and drugs to a schema based on the British National Formulary [9]. Strict quality control procedures, including duplicate data entry of a proportion of randomly selected records, were carried out throughout the study period. Ethical approval for the study was granted by the United Kingdom Multi-Regional Ethics Committee.

Primary care records were abstracted for 310 (97.5% of those interviewed) diffuse-large B-cell lymphoma (DLBCL) case/control matched pairs, 226 (99.1% of those interviewed) follicular lymphoma pairs (FL) and 225 (94.9% of those interviewed) Hodgkin lymphoma (HL) pairs. Matched case–control studies are often analysed using logistic regression conditional on the matched sets, using the case/control status as outcome and other variables thought relevant to the outcome as explanatory variables. However, in a 1-1 matched study, it is also possible to consider the case/control status as an explanatory variable in a regression that considers some other variable as outcome. This is because matching produces a case set and a control set that are nominally identical, as sets, with respect to the matching variables. Of course, the magnitude of any regression coefficients cannot be directly generalized from the sample to the population, but any qualitative difference between cases and controls remains valid.

In the present study, counts of visits to primary care (general practitioner) resulting in infectious disease diagnoses and non-infectious disease diagnoses per month were considered as separate longitudinal outcomes and modelled with negative binomial regression, using the number of months before lymphoma diagnosis (or pseudo-diagnosis), case control status and the interaction of these two variables as explanatory variables. As the counts of visits resulting in infectious diagnoses and in non-infectious diagnoses could now be considered as longitudinal outcomes, care was taken over the selection of the appropriate functional form for the time before lymphoma diagnosis/pseudo-diagnosis. In the models presented here, time before diagnosis was used untransformed. In addition, negative binomial generalized additive models (GAMs) were fitted [10] in order to investigate possible departures from these model assumptions. Where the results of generalized additive modelling depart from the main analysis, the differences are described below. In addition, each monthly count was treated as being independent from any other monthly count after diagnostic checks revealed evidence of only small levels of inter-monthly correlation. As a diagnostic check of this assumption, robust standard errors were calculated. In all cases these made negligible differences to the analysis. Confidence intervals based on robust standard errors are presented in Appendix A.

Inspection of the raw counts by month indicated a considerable inflation of diagnoses in the year before diagnosis/pseudo-diagnosis with lymphoma. In order to avoid the effects during this period from swamping effects earlier than this, the 12 months prior to diagnosis/pseudo-diagnosis were omitted from the models. All analyses were performed using STATA version 10.0 [11] and R version 2.9.2 [12] with the mgcv library used for the fitting of generalized additive models [10], the MASS library [13] for negative binomial regression and the sandwich library [14,15] for robust standard errors.

3. Results

Of the 761 cases with lymphoma, 310 had diffuse-large B-cell lymphoma (DLBCL), 226 had follicular lymphoma (FL) and 225 had Hodgkin lymphoma (HL). The median age at diagnosis and sex distribution for each type is shown in Table 1, together with the median number of visits for the different types of diagnoses made by the primary care physician (general practitioner) in the 15 years prior to diagnosis (including and excluding the year prior to diagnosis) for each lymphoma subtype and for controls. Overall, there were substantially more visits for non-infectious problems than for infections, both among cases and among controls.

Raw counts of visits to primary care (general practitioner) resulting in infectious and non-infectious diagnoses in the 15 years prior to diagnosis are shown for each subtype of lymphoma in Figs. 1 and 2 (cases are in red and controls in blue). For both types of primary care (general practitioner) diagnosis and for all subtypes of lymphoma, the counts rise markedly in the year prior to diagnosis. These data are modelled as described in Section 2 – excluding data from the year prior to diagnosis – and the linear relationship of non-infectious and infectious diagnoses in cases and controls is shown in Figs. 3 and 4 for each subtype of lymphoma. Model coefficients, 95%

Table 1
Q3 General practitioner (GP) visits for infectious and non-infectious diagnoses.

	Non-Hodgkin's lymphoma				Hodgkin's lymphoma	
	Diffuse-large B-cell		Follicular		Cases $N=225$	Controls $N=225$
	Cases $N=310$	Controls $N=310$	Cases $N=226$	Controls $N=226$		
Age at diagnosis/pseudo-diagnosis (median years)	54.4	54.4	54.1	54.1	38.8	38.8
Male (%)	167(53.9)	167 (53.9)	102 (45.1)	102(45.1)	142 (63.1)	142 (63.1)
GP visits in the 15 years before diagnosis/pseudo-diagnosis						
Infectious diagnosis						
Total visits	2561	2361	1872	1760	2390	1920
Median per person	6	5	6	6	8	6
Non-infectious diagnosis						
Total visits	19,535	17,387	16,839	14,729	12,037	10,589
Median per person	45	36	53.5	45	37	34
GP visits in the 15 years before diagnosis/pseudo-diagnosis (excluding visits in the year immediately before						
Infectious diagnosis						
Total visits	2228	2154	1712	1649	2194	1800
Median per person	5	5	5	5	7	6
Non-infectious diagnosis						
Total visits	16,236	15,825	14,444	13,327	10,126	9694
Median per person	37.5	33	44.5	42	31	30

Please cite this article in press as: Crouch S, et al. Illness patterns prior to diagnosis of lymphoma: Analysis of UK medical records. Cancer Epidemiology (2010), doi:10.1016/j.canep.2010.08.003

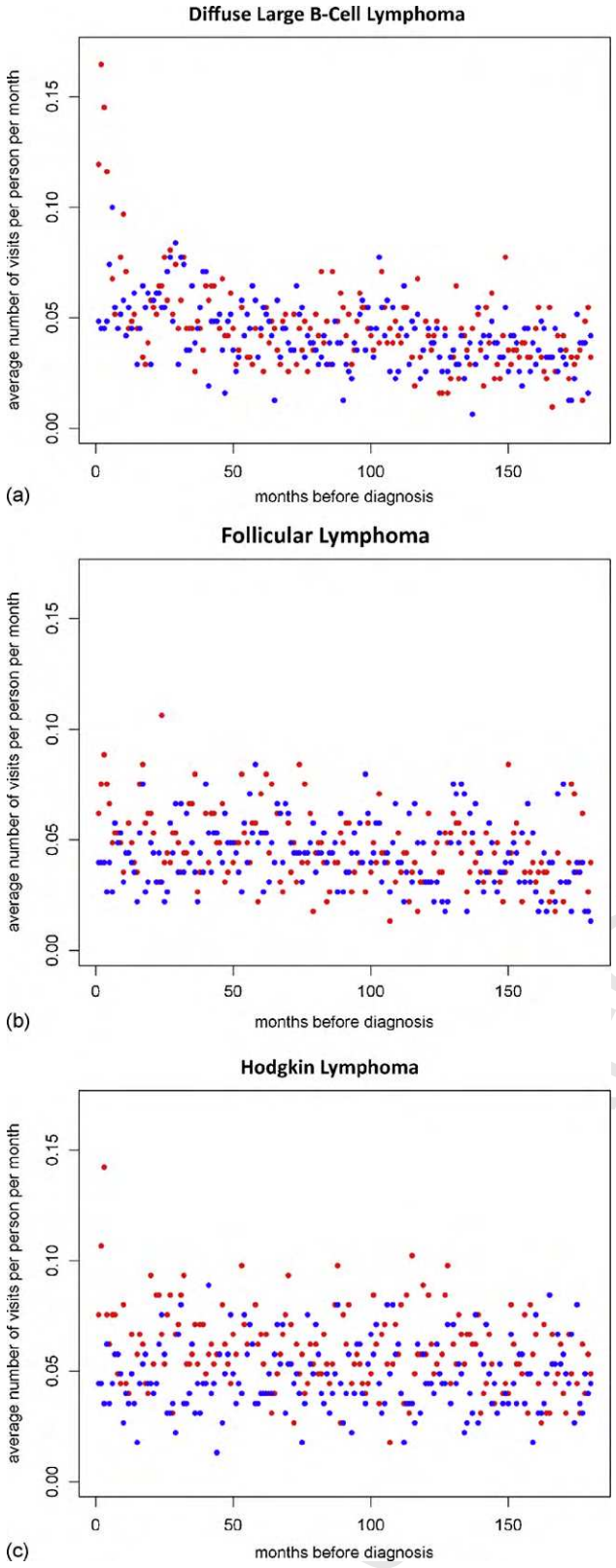

Fig. 1. (a) Raw counts of number of visits for diagnosed infections by month before diagnosis/pseudo-diagnosis for diffuse-large B-cell lymphoma. Cases are in red; Q2 controls in blue. (b) Raw counts of number of visits for diagnosed infections by month before diagnosis/pseudo-diagnosis for Follicular Lymphoma. Cases are in red; controls in blue. (c) Raw counts of number of visits for diagnosed infections by month before diagnosis/pseudo-diagnosis for Hodgkin lymphoma. Cases are in red; controls in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

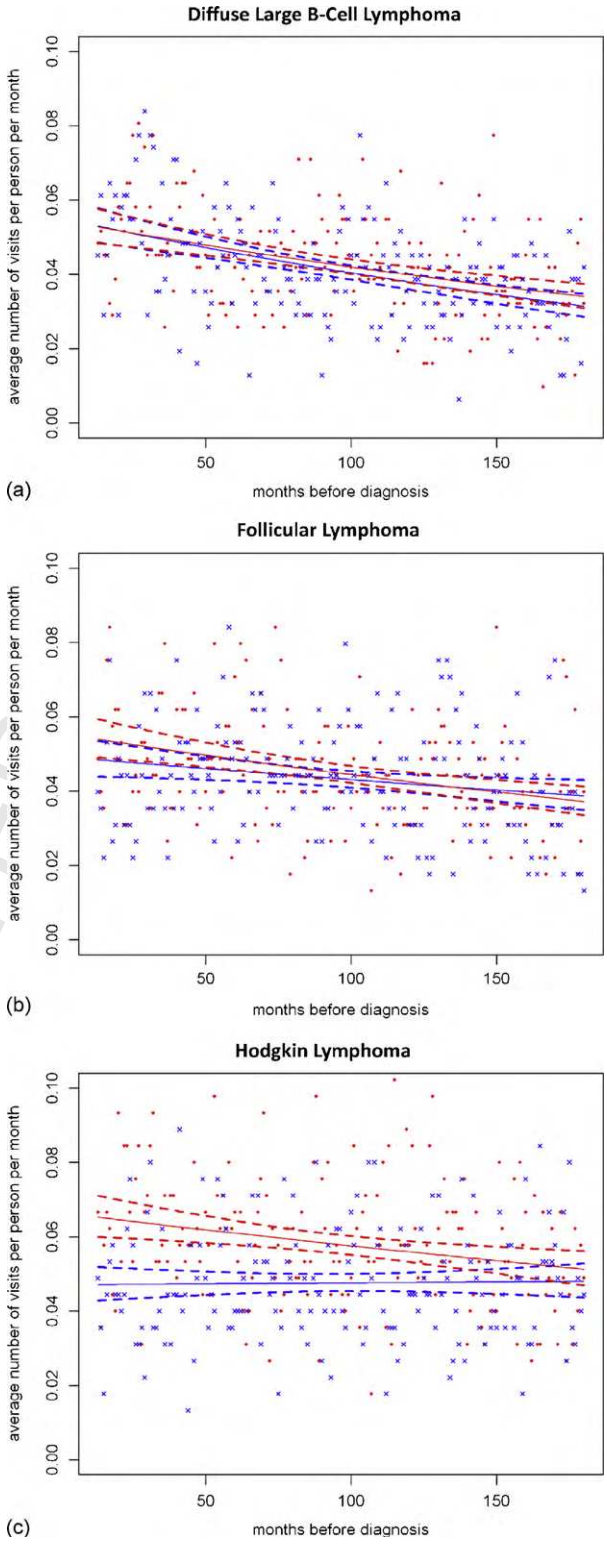

Fig. 2. (a) Fitted models for number of visits for infections from the 12 months before diagnosis/pseudo-diagnosis for diffuse-large B-cell lymphoma. Solid lines denote fitted models; dashed lines give 95% confidence intervals. Cases are in red; controls in blue. (b) Fitted models for number of visits for infections from the 12 months before diagnosis/pseudo-diagnosis for Follicular Lymphoma. Solid lines denote fitted models; dashed lines give 95% confidence intervals. Cases are in red; controls in blue. (c) Fitted models for number of visits for infections from the 12 months before diagnosis/pseudo-diagnosis for Hodgkin lymphoma. Solid lines denote fitted models; dashed lines give 95% confidence intervals. Cases are in red; controls in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

137

Please cite this article in press as: Crouch S, et al. Illness patterns prior to diagnosis of lymphoma: Analysis of UK medical records. Cancer Epidemiology (2010), doi:10.1016/j.canep.2010.08.003

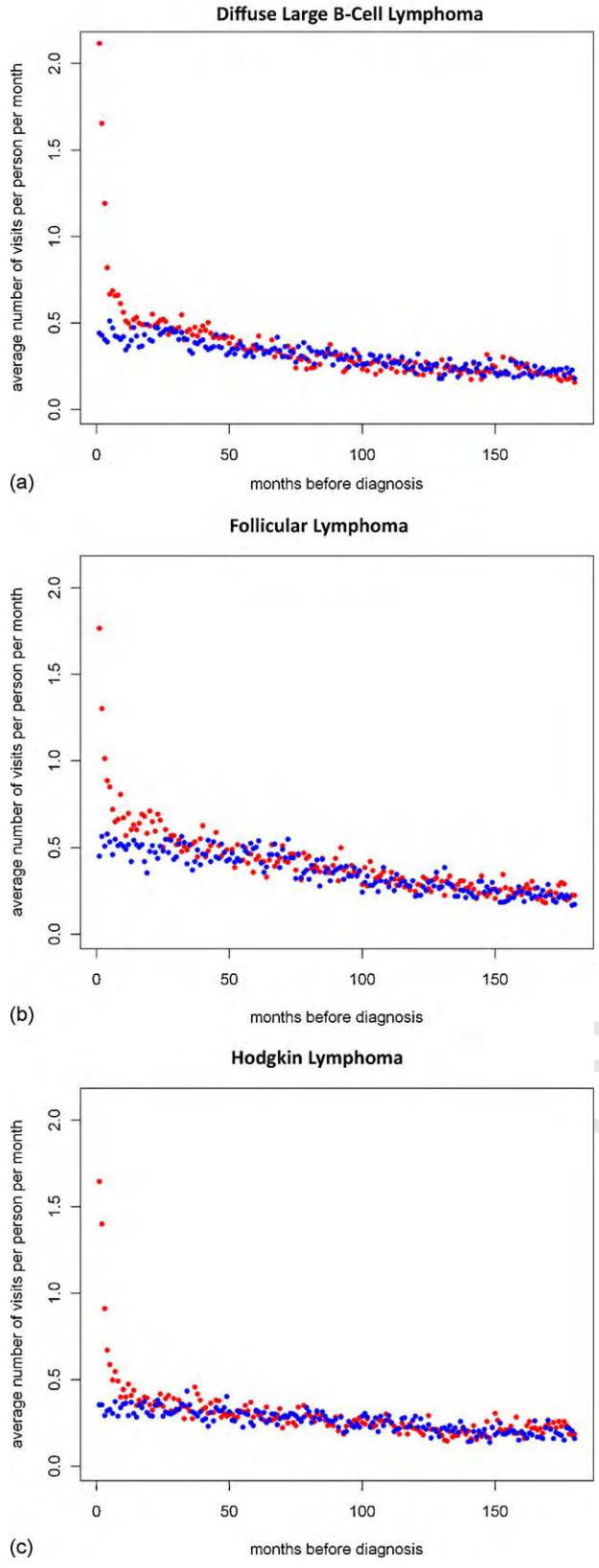

Fig. 3. (a) Raw counts of number of visits for diagnosed non-infections by month before diagnosis/pseudo-diagnosis for diffuse-large B-cell lymphoma. Cases are in red; controls in blue. (b) Raw counts of number of visits for diagnosed non-infections by month before diagnosis/pseudo-diagnosis for Follicular Lymphoma. Cases are in red; controls in blue. (c) Raw counts of number of visits for diagnosed non-infections by month before diagnosis/pseudo-diagnosis for Hodgkin lymphoma. Cases are in red; controls in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

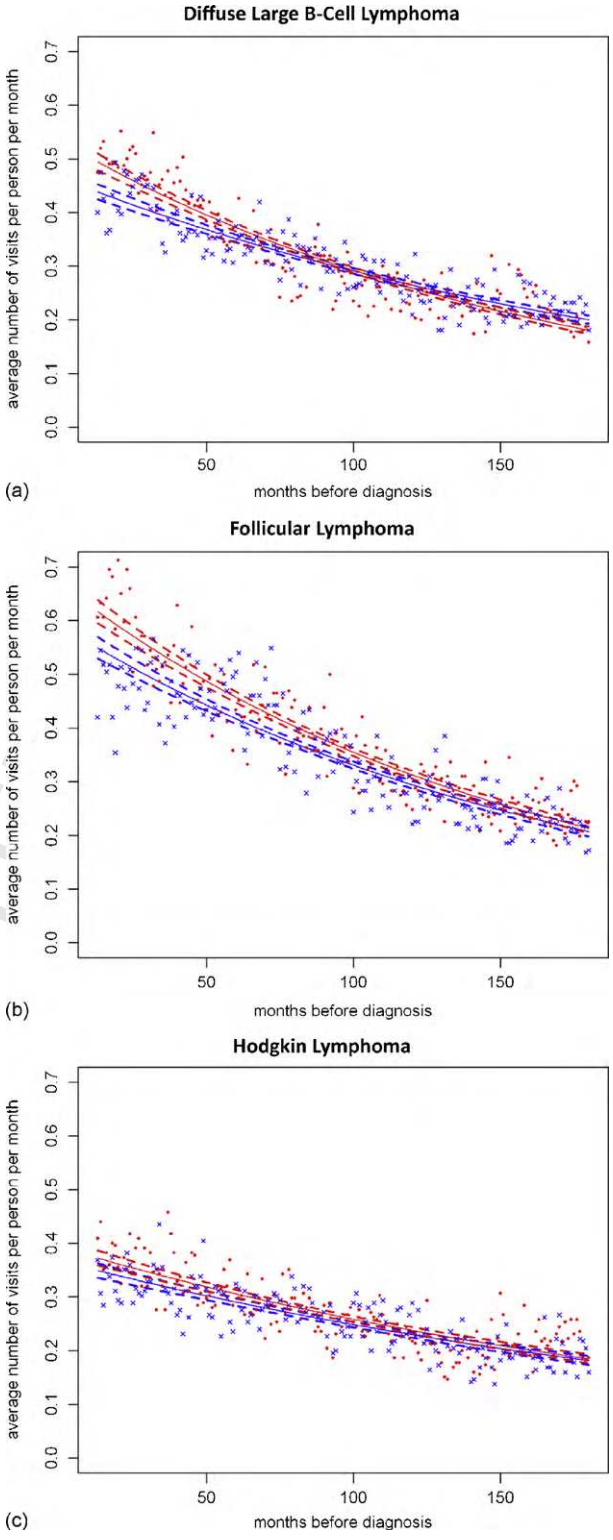

Fig. 4. (a) Fitted models for number of visits for non-infections from the 12 months before diagnosis/pseudo-diagnosis for diffuse-large B-cell lymphoma. Solid lines denote fitted models; dashed lines give 95% confidence intervals. Cases are in red; controls in blue. (b) Fitted models for number of visits for non-infections from the 12 months before diagnosis/pseudo-diagnosis for Follicular Lymphoma. Solid lines denote fitted models; dashed lines give 95% confidence intervals. Cases are in red; controls in blue. (c) Fitted models for number of visits for non-infections from the 12 months before diagnosis/pseudo-diagnosis for Hodgkin lymphoma. Solid lines denote fitted models; dashed lines give 95% confidence intervals. Cases are in red; controls in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Please cite this article in press as: Crouch S, et al. Illness patterns prior to diagnosis of lymphoma: Analysis of UK medical records. Cancer Epidemiology (2010), doi:10.1016/j.canep.2010.08.003

confidence intervals and corresponding p-values may be found in Appendix A.

Fig. 3 shows the fitted models for counts of visits resulting in infectious diagnoses among cases (the red line) and controls (the blue line) for the models linear in the time variable. There is no evidence of any difference between cases and controls in relation to the number or pattern of visits resulting in infectious diagnoses prior to the onset of DLBCL or FL. However, for HL, there is a clear excess of clinically diagnosed infections that is evident for at least a decade prior to lymphoma diagnosis. This case–control difference reflects a general increase in a broad range of infections, and is not due to any specific infection.

Fig. 4 shows the fitted models for counts of visits resulting in non-infectious diagnoses. Here there is clear evidence of a case–control difference for both DLBCL and FL; the excess being evident for between 4 and 6 years prior to lymphoma diagnosis. As for infections and HL, detailed examination of the records revealed that this association was non-specific in nature – with most visits being associated with symptoms such as tiredness, general malaise and depression. No differences for visits resulting in non-infectious diagnoses were evident for HL.

4. Discussion

Our results demonstrate substantial variation in the patterns of illness presenting to primary care physicians in the years preceding diagnosis of the lymphoma subtypes examined here. Excesses of visits resulting in infectious diagnoses were noted for HL and of visits resulting in non-infectious diagnoses for DLBCL and FL; in all cases the excesses were evident several years before lymphoma was diagnosed.

Although there are some differences in detail, the regression models with time before diagnosis untransformed and the GAMs reveal similar broad qualitative differences between the histories of visits for infectious diagnoses and for non-infectious diagnoses of these three conditions. As far as the history of visits for infectious diagnoses is concerned, there is little difference between cases and controls in DLBCL and FL, but in HL there is a marked divergence between cases and controls dating from as much as 10 years before diagnosis. For the history of visits for non-infectious diagnoses, patterns are more closely related. GAMs suggest divergence between cases and controls for DLBCL and FL between 4 and 6 years prior to diagnosis, with little difference between cases and controls prior to that divergence; no such effect was seen for HL. In summary, differences in patterns of attendance at primary care were evident between cases and controls (for years prior to diagnosis), but also between those with different types of lymphoma. The excess of visits resulting in infectious diagnoses prior to diagnosis of HL may suggest underlying immune abnormality, but we found little evidence of such an effect among patients subsequently diagnosed either with DLBCL or with FL. However, there is good evidence that infectious and inflammatory process may mediate risk of other lymphoma subtypes that were too rare to consider here, and larger population-based studies will be required [16–18].

Large amounts of information on previous illnesses, including infections, are routinely collected by medical practitioners working in primary care. Although these data, which are principally collected with the aim of documenting and monitoring patient care, have been used in a limited way in epidemiological studies their potential with respect to describing disease trajectories has yet to be fully realised [19–23]. A critical feature for aetiological and other studies – where the sequence and timing of events is important – is that information held in general practitioner records is collected prior to the diagnosis of malignancy and so has the advantage of being unaffected by recall and reporting bias [24].

Limitations of our study include its restricted age range (18–65 years), comparatively small size, and lack of information on other lymphoma subtypes [17]. With respect to the first of these, the median age at diagnosis of most lymphoproliferative malignancies exceeds 70 years, with the sex-specific rates varying with age (www.seer.cancer.gov; www.hmrn.org). DLBCL, for example, is more common in men, with the age-specific rates diverging as age increases; FL on the other hand is marginally more common in women with rates converging as age increases. By contrast, HL has a characteristic bimodal age distribution with a slight predominance of women at younger ages and of men at older ages – these patterns being reflective of different HL subtypes, which unfortunately we could not distinguish in the present dataset.

In conclusion, the different patterns of co-morbidity reported here, taken together with the different descriptive patterns, suggest different pathogenic mechanisms. Furthermore, the long prodromes suggested by our data indicate that disease may be present long before the diagnosis is made.

Conflict of interest

None.

Acknowledgement

Financial support for this work was provided by Leukaemia & Lymphoma Research (formerly the Leukaemia Research Fund).

Appendix A

Table A1.

Table A1
Model coefficients, confidence intervals and p-values.

	Coefficient	95% CI	p-Value
Infections			
DLBCL			
Intercept	2.84	(2.74, 2.94)	
Case/control	−0.0108	(−0.151, 0.129)	0.88
Months	-3.11×10^{-3}	$(-4.08\times10^{-3}, -2.14\times10^{-3})$	$<10^{-6}$
Interaction	4.97×10^{-4}	$(-8.66\times10^{-4}, 1.86\times10^{-3})$	0.47
Follicular			
Intercept	2.41	(2.30, 2.52)	
Case/control	0.119	(−0.0405, 0.278)	0.14
Months	-1.34×10^{-3}	$(-2.48\times10^{-3}, -2.03\times10^{-4})$	0.02
Interaction	-8.80×10^{-4}	$(-2.47\times10^{-3}, 7.06\times10^{-4})$	0.28
Hodgkin			
Intercept	2.36	(2.25, 2.47)	
Case/control	0.344	(0.203, 0.485)	1.72×10^{-6}
Months	1.06×10^{-4}	$(-9.13\times10^{-4}, 1.12\times10^{-3})$	0.84
Interaction	-1.54×10^{-3}	$(-2.86\times10^{-3}, -2.15\times10^{-4})$	0.023
Non-infections			
DLBCL			
Intercept	4.97	(4.94, 5.01)	
Case/control	0.138	(0.0895, 0.186)	$<10^{-6}$
Months	-4.69×10^{-3}	$(-5.05\times10^{-3}, -4.34\times10^{-3})$	$<10^{-6}$
Interaction	-1.34×10^{-3}	$(-1.85\times10^{-3}, -8.19\times10^{-4})$	$<10^{-6}$
Follicular			
Intercept	4.90	(4.85, 4.94)	
Case/control	0.123	(0.0637, 0.182)	4.5×10^{-5}
Months	-5.83×10^{-3}	$(-6.26\times10^{-3}, -5.40\times10^{-3})$	$<10^{-6}$
Interaction	-5.19×10^{-4}	$(-1.10\times10^{-3}, 5.84\times10^{-5})$	0.078
Hodgkin			
Intercept	4.41	(4.37, 4.46)	
Case/control	0.0683	$(8.3\times10^{-3}, 0.128)$	0.026
Months	-3.91×10^{-3}	$(-4.34\times10^{-3}, -3.49\times10^{-3})$	$<10^{-6}$
Interaction	-2.85×10^{-4}	$(-9.33\times10^{-4}, 3.63\times10^{-4})$	0.39

Please cite this article in press as: Crouch S, et al. Illness patterns prior to diagnosis of lymphoma: Analysis of UK medical records. Cancer Epidemiology (2010), doi:10.1016/j.canep.2010.08.003

References

[1] Smedby KE, Askling J, Mariette X, Baecklund E. Autoimmune and inflammatory disorders and risk of malignant lymphomas—an update. J Intern Med 2008;264(6):514–27.
[2] Grulich AE, van Leeuwen MT, Falster MO, Vajdic CM. Incidence of cancers in people with HIV/AIDS compared with immunosuppressed transplant recipients: a meta-analysis. Lancet 2007;370(9581):59–67.
[3] Bouvard V, Baan R, Straif K, et al. A review of human carcinogens—Part B: biological agents. Lancet Oncol 2009;10(4):321–2.
[4] Newton R, Crouch S, Ansell P, et al. Hodgkin's lymphoma and infection: findings from a UK case–control study. Br J Cancer 2007;97(9):1310–4.
[5] Willett EV, Smith AG, Dovey GJ, et al. Tobacco and alcohol consumption and the risk of non-Hodgkin lymphoma. Cancer Causes Control 2004;15(8):771–80.
[6] World Health Organization. International classification of disease for oncology, third ed., Geneva: World Health Organisation, 2000.
[7] Willett EV, Morton LM, Hartge P, et al. Non-Hodgkin lymphoma and obesity: a pooled analysis from the InterLymph Consortium. Int J Cancer 2008;122(9):2062–70.
[8] World Health Organization. Tenth revision of the international statistical classification of diseases and Related Health Problems (ICD-10). World Health Organization; 1992.
[9] British Medical Association RPSoGB. British National Formulary—Number 24 (September). London: British Medical Association and The Pharmaceutical Press, 1992.
[10] Wood SN. Generlized additive models: an introduction. Chapman Hall; 2006.
[11] StataCorp LP. Stata/IC 10.1 for Windows. College Station, TX: StataCorp LP, 2007.
[12] R Development core team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for statistical computing, 2009.
[13] Venables WN, Ripley BD. Modern applied statistics with S, fourth ed., New York: Springer, 2002.
[14] Zeileis A. Econometric computing with HC and HAC covariance matrix estimators. Journal of Statistical Software 2004;11(10):1–17.
[15] Zeileis A. Object-orientated computation of sandwich estimators. Journal of Statistical Software 2006;16(9):1–16.
[16] Smith A, Roman E, Howell D, et al. The Haematological Malignancy Research Network (HMRN): a new information strategy for population based epidemiology and health service research. Br J Haematol 2009.
[17] World Health Organization. WHO classification of tumours of haematopoietic and lymphoid tissues. Lyon: International Agency for Research on Cancer, 2008
[18] Sehn LH, Donaldson J, Chhanabhai M, et al. Introduction of combined CHOP plus rituximab therapy dramatically improved outcome of diffuse large B-cell lymphoma in British Columbia. J Clin Oncol 2005;23(22):5027–33.
[19] Ansell P, Johnston T, Simpson J, et al. Brain tumor signs and symptoms: analysis of primary health care records from the UKCCS. Pediatrics 2010;125(1):112–9.
[20] Cardwell CR, McKinney PA, Patterson CC, Murray LJ. Infections in early life and childhood leukaemia risk: a UK case–control study of general practitioner records. Br J Cancer 2008;99(9):1529–33.
[21] Chilvers CE, Pike MC, Taylor CN, et al. General practitioner notes as a source of information for case–control studies in young women. UK National Case–Control Study Group. J Epidemiol Community Health 1994;48(1):92–7.
[22] Howell DA, Smith AG, Roman E. Lymphoma: variations in time to diagnosis and treatment. Eur J Cancer Care (Engl) 2006;15(3):272–8.
[23] Roman E, Simpson J, Ansell P, et al. Childhood acute lymphoblastic leukemia and infections in the first year of life: a report from the United Kingdom Childhood Cancer Study. Am J Epidemiol 2007;165(5):496–504.
[24] Simpson J, Smith A, Ansell P, Roman E. Childhood leukaemia and infectious exposure: a report from the United Kingdom Childhood Cancer Study (UKCCS). Eur J Cancer 2007;43(16):2396–403.

Please cite this article in press as: Crouch S, et al. Illness patterns prior to diagnosis of lymphoma: Analysis of UK medical records. Cancer Epidemiology (2010), doi:10.1016/j.canep.2010.08.003