



HHS Public Access

Author manuscript

Br J Haematol. Author manuscript; available in PMC 2019 May 01.

Published in final edited form as:

Br J Haematol. 2018 May ; 181(4): 505–514. doi:10.1111/bjh.15229.

Risk factors for Burkitt lymphoma: A nested case-control study in the UK Clinical Practice Research Datalink

Parisa Karimi¹, Brenda M. Birmann², Lesley A. Anderson³, Charlene M. McShane³, Shahinaz M. Gadalla⁴, Joshua N. Sampson⁵, and Sam M. Mbulaiteye¹

¹Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, 9609 Medical Ctr Dr, Bethesda, MD, 20892 USA

²Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 181 Longwood Avenue, Boston, MA, 02115 USA

³Centre for Public Health, School of Medicine, Dentistry and Biomedical Science, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom

⁴Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, 9609 Medical Ctr Dr, Bethesda, MD, 20892 USA

⁵Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, 9609 Medical Ctr Dr, Bethesda, MD, 20892 USA

Summary

Burkitt lymphoma (BL) occurs as three subtypes: endemic BL, immunosuppression-related BL and sporadic BL. Descriptive studies of BL age-specific incidence patterns have suggested multimodal peaks near 10, 40 and 70 years of age, but the risk factors for BL at different ages are unknown. We investigated risk factors for BL in the United Kingdom among 156 BL cases and 608 matched BL-free controls identified in the Clinical Practice Research Datalink (CPRD) between 1992 and 2016. Associations with pre-diagnostic body mass index, cigarette smoking, alcohol consumption, hepatitis, Epstein–Barr virus (EBV), human immunodeficiency virus infection and acquired immune deficiency syndrome (HIV/AIDS), malaria, allergic and autoimmune conditions, and prednisone use were evaluated. Overall, we identified inverse associations between smoking and BL risk, and positive associations between prior EBV infection, HIV/AIDS and prescription or use of prednisone with BL risk. In age-group stratified analyses, BL was associated with malaria exposure (vs. no exposure, Odds ratio [OR] 8.00, 95% confidence

Correspondence: Sam M. Mbulaiteye, Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, 9609 Medical Ctr Dr, Bethesda, MD, 20892 USA. Tel: +1 240 276 7108; Fax: +1 240 276 7836; mbulais@mail.nih.gov.

DR. CHARLENE MCSHANE (Orcid ID: 0000-0001-8609-0788)

DR. SAM M MBULAITEYE (Orcid ID: 0000-0002-8273-9831)

Authorship contributions

Sam M. Mbulaiteye conceived and designed the study. Parisa Karimi and Sam M. Mbulaiteye analysed and interpreted the data and drafted the paper. Brenda M Birmann, Lesley A. Anderson, Charlene M. McShane, Shahinaz M. Gadalla, and Joshua N. Sampson interpreted data and edited the paper. All authors critically reviewed and approved the final manuscript.

Conflicts of interest

The authors declared no competing interest

interval [CI] 1.46 – 43.7) among those aged 20–59 years old and with hepatitis infection (vs. no infection, OR 3.41, 95%CI 1.01–11.5) among those aged 60+ years old. The effects of EBV, malaria, HIV/AIDS, prednisone and hepatitis on BL remained significant in mutually-adjusted age-group-specific analyses. No risk factors were associated with childhood BL. We report novel associations for BL in non-endemic settings.

Keywords

Epstein–Barr virus; Falciparum malaria; Burkitt lymphoma; prednisone; smoking

Introduction

Burkitt lymphoma (BL) is an aggressive B cell lymphoma that occurs as three subtypes: endemic BL, immunosuppression-related BL, and sporadic BL (Leoncini, *et al* 2008). Endemic BL (eBL) is the most common paediatric cancer in many countries in sub-Saharan Africa (Ogwang, *et al* 2008, Parkin, *et al* 2008), and classically presents as a facial or abdominal tumour in children aged 6–9 years. Immunosuppression-related BL occurs at a much higher incidence in immunosuppressed (20- to 60-fold vs. immune competent) persons, such as those with human immunodeficiency virus (HIV) infection (Guech-Ongey, *et al* 2010, Mbulaiteye, *et al* 2003), receiving immunosuppressive drugs after solid organ transplantation (Mbulaiteye, *et al* 2013), or with rare congenital immunodeficiency syndromes, such as XMEN (X-linked immunodeficiency with magnesium defect, Epstein-Barr virus (EBV) infection and neoplasia) (Li, *et al* 2014), and classically presents as nodal disease. Sporadic BL (sBL) occurs worldwide in people living in non-malaria endemic regions, such as the United States (US) and the United Kingdom (UK), and often presents as nodal or leukaemic disease (Levine, *et al* 1982). EBV(de-The, *et al* 1978) and *Plasmodium falciparum* malaria(Aka, *et al* 2013, Carpenter, *et al* 2008, Legason, *et al* 2017) have been identified as risk factors for eBL, but the risk factors for sBL are unknown.

Recent reports of steep annual increases in the age-standardized incidence rates of BL, including increases of 6–7% per year between 1973 and 2005 in the US (Mbulaiteye, *et al* 2010) and of 5–6% per year between 1963 and 2002 in four continents (excluding Africa) (Mbulaiteye, *et al* 2012), have rekindled an interest in identifying risk factors for BL in low BL incidence countries. Although the annual increase in the age-standardized incidence rates of BL could partly be due to the contribution from HIV-related BL cases, these increases began before the onset of the HIV epidemic and persisted after the introduction of life-saving treatments for HIV (Mbulaiteye, *et al* 2010; Mbulaiteye, *et al* 2012). Furthermore, the increases were observed in analyses restricted to registries excluding those US regions with many acquired immune deficiency syndrome (AIDS) cases, such as New Jersey, New York and California, or when analyses were repeated after reducing the number of BL cases by about 25% (Mbulaiteye, *et al* 2010), the proportion attributable to HIV infection. Another reason for interest in BL risk factors stems from findings in descriptive studies of BL that have shown multiple age-specific incidence peaks near 10, 40 and 70 years of age (Mbulaiteye, *et al* 2010; Mbulaiteye, *et al* 2012), suggesting aetiological heterogeneity for BL in age-defined populations (Macmahon 1957). Interestingly, a

molecular study of BL conducted recently reported the presence of different pattern of somatic mutations in BL tumours diagnosed at different ages (Havelange, *et al* 2016, Mbulaiteye and Anderson 2018). Furthermore, an analysis of epidemiological risk factors for sBL cases in the International Lymphoma Epidemiology Consortium (InterLymph) suggested differences in the risk factors for sBL in persons aged <50 years versus ≥50 years old (Mbulaiteye, *et al* 2014a). In the current study, we used longitudinal data from the Clinical Practice Research Datalink (CPRD) to investigate risk factors for BL in the United Kingdom (UK), one of the countries that contributed data to the four-continent BL study that showed steep annual increases in the age-standardized incidence rates of BL (Mbulaiteye, *et al* 2012).

Methods

Description of the CPRD population and data

The CPRD is a longitudinal anonymized database established in 1987 to collect electronic primary care medical records from more than 674 general practices (Herrett, *et al* 2015). The CPRD electronically captures routine demographic, clinical and laboratory data collected by general practitioners for more than 22 million patients, of whom over 4.4 million are currently alive and registered. The CPRD data have been shown to be demographically and sociologically representative of the UK population in behavioural characteristics, such as smoking patterns (Booth, *et al* 2013), and the accuracy and completeness of recording of certain medical diagnoses, including haematological malignancies, has been verified (Dregan, *et al* 2012). The data quality for participating practices are evaluated and assigned an “Up-to-Standard” (UTS) date, up to which the given clinic’s data are considered to be of research quality (Herrett, *et al* 2015).

Ethical approval—All data were anonymized before creating analysis datasets to avoid the possibility of patient identification. Ethical approval to conduct the study was obtained from the Office of Human Subjects Research at the National Institutes of Health. The UK Independent Scientific Advisory Committee (ISAC) at CPRD approved the study (Protocol number: ISAC-15_135R).

Eligibility, definition, and classification of BL cases and controls

We used CPRD GOLD July 2016 data release. The eligible participants included those registered between 1992 and 2016 who had at least two years of follow-up in the CPRD before their index date (defined below). To maximize the quality of data used in the analysis, we further restricted the study to participants followed in clinics with at least two years of UTS follow-up (Figure 1). The CPRD-eligibility criteria included having a minimum follow-up of two years before the index date, defined as the date of BL diagnosis minus one year, and being in a clinic with at least two years of UTS follow-up.

The BL cases in CPRD were identified using “Read codes” for a BL diagnosis in the cohort that met CPRD-related eligibility criteria defined above. The BL Read codes included A78960 (HIV-related), B602100, B602500, B602200, B602300 (BL in the lymph nodes), BBr2600 (bone-marrow or leukaemia BL), and B602.00, B602z0 (unspecified site). As no

cases of transplant-related BL were identified, we considered all but the HIV-related cases to be sBL.

The controls were BL-free individuals at their index date who were sampled using risk-set or density sampling, i.e., controls entered before their case's entry and exit after the case's exit. Controls were assigned the index date of their matched case. Three to four controls per BL case were sampled and matched to the case on age, sex, race, general practice or region (when appropriate controls could not be matched on the general practice). Because race was missing for many subjects, cases with missing race were matched to controls with missing race.

Risk factors

To minimize the risk of reverse causation, we used risk factor data collected at least two years before the index date for both cases and controls. The investigated potential risk factors included *a priori* selected lifestyle and behavioural characteristics (body mass index [BMI], smoking and alcohol status), and aspects of the medical history of clinical events (any infection, hepatitis, EBV, HIV/AIDS, malaria, allergy, hay fever, asthma, eczema, gout, and ever use of prednisone)(Mbulaiteye, *et al* 2014a). Read codes were used to identify and define the potential risk factors (Wang, *et al* 2018). A variable named "any infection" was created based on having a Read code for specified or unspecified infections. The specified infections include those due to viruses (e.g., HIV, EBV, hepatitis A, B or C viruses), bacteria (e.g., *streptococcal*, *E. coli*, *Klebsiella*, *Haemophilus influenza*, *mycoplasma* or *salmonella*), or parasites (malaria). The unspecified infectious included syndromes caused by infection, such as septicaemia, cellulitis, pneumonia or pyelonephritis. This variable was designed to assess the likelihood of BL cases being prone to specific or non-specific infections, which might suggest a correlation between prior infection and risk for BL. Read codes for EBV included those for a diagnosis of EBV or a record of positive serology for EBV; Read codes for exposure to malaria included those for malaria treatment, reported a history of malaria, or receipt of malarial prophylaxis. Because this study is exploratory, most risk factors were defined as ever versus never; a classification of "ever" indicated the finding of at least one Read code consistent with a positive result for the risk factor. A classification of "never" indicated either the absence of a code suggesting exposure or the presence of an explicit code for non-exposure. BMI was calculated from height and weight values, and categorized into tertiles, based on cut-off values calculated using overall or age-specific data from the controls.

Statistical analysis

We calculated Spearman partial correlation coefficients, with adjustment for age and sex, to examine the pairwise correlations of the main risk factors of interest. Using conditional logistic regression for matched sets, we estimated odds ratios (ORs) and 95% confidence intervals (95% CIs) for the association of BL with each independent variable. Missing data were coded as a specific category, but individuals with missing data for the exposure variable of interest were excluded from the model when estimating the association. To explore the potential aetiological heterogeneity of age-defined BL (Mbulaiteye, *et al* 2010) and assess whether regional factors influence BL risk, we conducted additional analyses

stratified by age groups (<20, 20–59 and 60+ years old) and by region (three regions with 40% of cases *versus* the ten regions with 60% of cases). We performed multivariate logistic regression age-specifically, including covariates associated with BL at a $P < 0.10$ and those considered *a priori* as risk factors (malaria exposure). Variables that did not improve the model fit, based on comparing the restricted and unrestricted models using the likelihood-ratio test, were removed from the final multivariate models. Because HIV/AIDS is a risk factor for BL, we repeated analyses after excluding HIV positive cases and controls, and the associations that persisted are interpreted as associations with sBL. After noting unexpected associations for smoking (see Results), we reviewed apparent never-smokers' entire CPRD record to evaluate the reliability of their never smoker status. Because our study was conducted to generate hypotheses about risk factors for BL in low endemicity settings, we did not adjust for multiple comparisons. The distribution of the cases by region was age-standardized by the direct method to the UK 2012 census population, assuming the cases arose from a population mirroring the CPRD active population in 2013. A two-sided $P < 0.05$ was considered statistically significant. Statistical analyses were conducted with Stata version 14 (Stata Corp, College Station, Texas).

Results

A total of 156 BL cases and 608 controls were studied. By design, the distributions of sex and age at CPRD registration and index date, and region of residence were similar for cases and controls (Table I). Most participants were aged 20–59 years or 60+ years; few were aged <20 years. The average number of years of observation in the CPRD was high, but lower in the BL cases than controls (17.4, standard deviation [SD]: 15.3) *vs.* controls (20.9 years, SD: 13.9; $P = 0.005$). About 40% of the BL cases were from three regions (West Midlands, South Central and Scotland) and 60% of the cases were from the remaining ten regions, including London, which is the most populous region. There were 19 (12.2%) BL cases, and four (0.7%) controls who were positive for HIV/AIDS in this sample. Ten (52.6%) of the BL cases and three (75.0%) of the controls with HIV/AIDS were from the three regions with 40% of the cases. We identified evidence for malaria exposure in 15 participants. The Read codes for malaria exposure in these patients indicated receipt of malaria prophylaxis in eight participants, a history of malaria in two participants, and a malaria screening test or malaria diagnosis in each of participants.

In the logistic regression analyses of individual risk factors (Table I), compared to never smokers, the risk of BL was significantly decreased in former smokers (OR 0.47, 95% CI 0.29 – 0.76) and in those who were current smokers (OR 0.24, 95% CI 0.12 – 0.46; P for trend < 0.0001). We confirmed that all the participants classified as never-smokers were consistently classified as never smokers before and after their index date, except 8 participants (all controls) who were classified as current or former smokers in at least one record after their index date. The association of smoking status with risk for BL did not change when the status of these eight controls was considered as misclassified and changed to a current or former smoker (data not shown). Compared to current or former drinkers, the risk of BL was non-significantly increased in never drinkers (OR 1.43, 95% CI 0.75 – 2.72). The association of smoking with BL was not explained by alcohol consumption because the variables were not significantly correlated with among the controls ($\rho = 0.06$, $P = 1$). The

associations of current (OR 0.25, 95% CI 0.13 – 0.49) and former smoking (OR 0.45, 95% CI 0.27 – 0.75) with BL risk did not change with adjustment for alcohol consumption. BMI was unrelated to the risk of BL (Table I).

With respect to infections, BL risk was elevated in participants with a record indicating hepatitis infection (hepatitis infected vs. uninfected: OR 1.88, 95% CI 1.02 – 3.47; Table II). Risk was also elevated in participants with a record indicating EBV infection (EBV-infected vs. uninfected: OR 4.50, 95% CI 1.74 – 11.7; Table II). Of interest, 15 of the 17 participants with a record of EBV infection also had a history of infectious mononucleosis, a condition that is a manifestation of delayed first exposure to the EBV and that may result in longer-term severe or chronic EBV. BL risk was not associated with hepatitis or immune-related conditions (allergy, hay fever, asthma, eczema or gout; Table II).

The risk for BL was increased in persons with HIV/AIDS (vs. no history of HIV/AIDS, OR 24.5, 95% CI 7.25 – 83.0; Table II). BL risk was also associated, albeit non-significantly, with having some evidence of exposure to malaria (vs. no exposure, OR 2.00, 95% CI 0.68 – 5.85). The risk for BL was associated with a history of prednisone use (vs. never used, OR 3.38, 95% CI 1.30 – 8.75). After excluding HIV/AIDS positive cases and controls, the associations between EBV, malaria and BL were only slightly attenuated, while the associations with prednisone became slightly stronger (Table S1).

The age-standardized BL prevalence of BL in the three regions with 40% of cases was 4.5 per 100,000 in comparison to 3.5 per 100,000 in the ten regions that contributed 60% of cases. Analysis of risk factors for BL stratified by region revealed similar patterns to the unstratified findings summarized above, except that the stratified analyses suggested a more pronounced association of BL risk with malaria in the three regions with 40% cases than in the ten regions with 60% of the cases (OR 3.15, 95% CI 0.62 – 16.1 and OR 1.21, 95% CI 0.24 – 6.09, respectively, $p=0.81$). The association of BL risk with HIV/AIDS was attenuated in the three regions with 40% of the cases and more pronounced in the ten regions with 60% of the cases (OR 12.1, 95% CI 2.51 – 57.9 *versus* OR 35.8, 95% CI 4.58 – 280, $p=0.001$), but with wide confidence intervals.

The age-stratified analysis excluded participants aged <20 years old because of sparse data. Among those aged 20–59 years old, compared to being a non-smoker, the risk of BL was significantly decreased among former smokers (OR 0.25, 95% CI 0.12 – 0.50) and current smokers (OR 0.20, 95% CI 0.09 – 0.46; P trend <.0001) and elevated in those with evidence of EBV infection (OR 5.60, 95% CI 1.78 – 17.6), HIV/AIDS (OR 20.6, 95% CI 5.98 to 70.7), or malaria exposure (OR 8.00, 95% CI 1.46 to 43.7). Among those aged 60+ years old, compared to being a non-smoker, the risk of BL was non-significantly decreased among former smokers (OR 0.76, 95% CI 0.36 – 1.58) and significantly decreased among current smokers (OR 0.18, 95% CI 0.05 – 0.67; P trend =0.014). The risk was marginally elevated in those with evidence of hepatitis infection (OR 3.41, 95% CI 1.01 – 11.5), elevated, but not significantly, in those with EBV infection (OR 4.00, 95% CI 0.56 to 28.4), and significantly elevated among those with a history of prednisone use (OR 7.30, 95% CI 1.81 – 29.4).

Multivariate analyses were performed separately for those aged 20–59 years and 60+ years old (Table III), based on our causal model that age might confound certain risk factors for BL (Mbulaiteye and Anderson 2018). The significant risk factors of BL among those aged 20–59 years old included former (OR 0.22, 95% CI 0.09 – 0.52) and current smoking (OR 0.12, 95% CI 0.04 – 0.35), EBV infection (OR 5.69, 95% CI 1.29 – 25.1), HIV/AIDS (OR 40.4, 95% CI 7.74 – 211.1) and malaria infection (OR 44.0, 95% CI 2.63 – 75.3). Among those aged 60+ years old, BL risk was significantly decreased among current smokers (OR 0.14, 95% CI 0.03 – 0.58) and significantly increased among those with a history of prednisone use (OR 7.78, 95% CI 1.85 – 32.5), and marginally elevated in those with hepatitis (OR 3.21, 95% CI 0.89 – 11.6) and EBV infection (OR 11.8, 95% CI 0.96 – 144.2). Restricting the adjusted models to HIV/AIDS negative participants did not materially change the EBV and prednisone associations, but associations with malaria were attenuated (Table S2).

Discussion

We present the first population-based study of risk factors for BL in the UK conducted in a longitudinal dataset obtained from the CPRD. Overall, we observed inverse associations between smoking and BL risk, and positive associations between prior EBV infection, HIV/AIDS and prescription or use of prednisone with BL risk. In age-group stratified analyses, BL risk was also significantly elevated among those aged 20–59 years old with a history of exposure to malaria, whereas a history of hepatitis use was associated with elevated risk of BL among those aged 60+ years old.

EBV is risk factor for eBL (Rochford, *et al* 2005), but our study presents data for novel association with BL among people in the UK aged 20–59 years and 60+ years old. We noted that EBV in this population appears to capture a history of infectious mononucleosis because 15 of 17 participants with a CPRD record of EBV infection also had a history of infectious mononucleosis, which is a more severe form of EBV infection (Klein, *et al* 1968) and a known risk factor for EBV-positive Hodgkin lymphoma (Hjalgrim, *et al* 2003). Despite the small numbers of participants with EBV codes in our CPRD sample, the associations with EBV were robust in multivariate age-stratified analyses, as well as in analyses excluding HIV/AIDS positive participants. The persistence of the associations after HIV-positive subjects were excluded suggests that the associations are for sBL, and that EBV infection, which is detected in up to 22% of sBL tumours using fluorescent in-situ hybridization studies (Mbulaiteye, *et al* 2014b), influences the risk for a small proportion of sBL in the UK. Our results encourage the design hypothesis-driven studies in large cohorts, such as the UK Millennium Cohort (Gares, *et al* 2017), to confirm or refute the hypothesized role of EBV infection in sBL.

Our results for malaria should be cautiously interpreted because they are partially based on Read codes suggesting people received prophylaxis to protect against malaria infection during international travel, and not actual infection. Because the postulated malaria exposure occurred 2–10 years before BL diagnosis, the results cannot be explained by reverse causality. The results suggest that short-term exposure to malaria parasites, even when moderated by medical suppression of infection, may influence the development of eBL 2–10

years later. A role of short-term exposure to malaria contrasts with the conventional thinking that BL risk is only related to chronic recurrent malaria occurring over years (Aka, *et al* 2013), and warrants urgent confirmation because hundreds of thousands of people from the UK travel to malaria-endemic areas from non-malaria endemic countries every year (Tatem, *et al* 2017). It is estimated that about 1,600 to 2,500 of these travellers suffer incidental clinical malaria and 9–15 suffer malaria-related deaths. If our results are confirmed, travellers from malaria-free countries to malaria endemic countries, who already are monitored for imported malaria (Schlagenhauf, *et al* 2015, Tatem, *et al* 2017, Warne, *et al* 2014), could also be monitored for BL. We note that the number of incidental malaria cases diagnosed between 1973 and 2014 among US citizens traveling to malaria-endemic countries increased 10-fold (Mace and Arguin 2017). During the same period, the age-standardized incidence rates of BL increased 6–7% per year (Mbulaiteye, *et al* 2010). Based on our results, we speculate that the steep increases in the age-standardized incidence rates of BL that have been reported in developed countries (Mbulaiteye, *et al* 2010, Mbulaiteye, *et al* 2012) could be related to exposure to malaria during international travel.

Our finding of a positive association between prior prescription or use of prednisone and risk of BL in participants aged 60+ years old may be due to chance. It could also be explained by the immunosuppressive effects of corticosteroids (Coutinho and Chapman 2011), or the association with underlying medical conditions for which the prednisone was prescribed. However, this explanation is not supported by the null association between BL and allergic conditions or asthma, some of the conditions for which prednisone may be prescribed. If the association with prednisone is confirmed, it would support careful monitoring of persons aged 60+ years old being treated with prednisone for complication with BL.

Our findings of strong, independent inverse associations between smoking and the risk for BL were unexpected. Although no association between smoking and BL risk was observed in the InterLymph pooled study of BL (Mbulaiteye, *et al* 2014a), which included 295 cases from multiple countries and had data available for some dose-response-type analyses, the associations in the current study were robust in several analyses and were not explained by misclassification of smoking status. The discrepancies between the present and the InterLymph studies may be explained by differences in how smoking was ascertained. Neither study had adequate statistical power to examine more detailed smoking histories to assess dose-response, especially in age-stratified analyses. We considered that biased selection of controls might have contributed to the present study findings; however, the distribution of smoking in our controls is similar to that observed in a much larger cohort of 45,000 apparently healthy participants in CPRD (Kuo, *et al* 2016). We note that inverse associations have been reported between smoking and risk of Kaposi sarcoma (KS) (Anderson, *et al* 2008, Goedert, *et al* 2002) (Nawar, *et al* 2005), which is caused by human herpesvirus 8 (Chang, *et al* 1994). The smoking-KS findings remain unexplained, and their relevance to our BL results is unclear; nonetheless, if confirmed, the BL and KS data collectively suggest that chemicals in cigarette smoke or other factors correlated with cigarette smoking may be associated with protection against two malignancies that are associated with two gamma-herpesviruses.

An investigation of risk factors for BL in the UK is difficult because it is rare. Thus, our results from CPRD should be considered exploratory and useful for generating hypotheses that may be evaluated in pooled studies. The strengths of our study include access to a large, representative longitudinal database and utilizing a prospective design. The limitations include incomplete data, such as race, which limits our ability to make inferences about the proportion of BL in different races, and whether BL cases were more likely to occur in immigrants from endemic malaria areas. Although this is theoretically possible, we believe this is unlikely because there are no published reports suggesting that immigrants to the UK from malaria endemic countries have a high incidence of BL. Furthermore, the average period of observation in CPRD in both cases and controls was high, which suggests that the cases and controls were more likely settled UK citizens. The small sample size resulted in imprecise estimates and limited our statistical power to observe weak or modest associations, especially for stratified analyses. We note that the presentation of both marginal and age-stratified results presents a dilemma, also called Simpson's paradox (Hernan, *et al* 2011), when the results for some variables, such as malaria or prednisone, do not agree and force the question about which estimates to believe when interpreting the data. We resolved this paradox by relying on the causal context of our investigation, which considered age as a confounder (Mbulaiteye, *et al* 2010, Mbulaiteye, *et al* 2012). Plausibly, age is an important modifier of the likelihood of international travel and, therefore, the risk of being exposed to malaria (Schlagenhauf, *et al* 2015, Tatem, *et al* 2017, Warne, *et al* 2014). Similar reasons apply to the likelihood of being given prednisone as a treatment. Based on these considerations, the marginal ORs of association for malaria and prednisone are confounded, which the ORs conditional on age reflect the contextually relevant effects, as discussed above. Our study relied on data collected during routine medical encounters, which are vulnerable to biases resulting from incompleteness or inaccuracy of information, and to temporal trend biases. No risk factors were identified for BL in children and adolescents, perhaps because the collected data were not appropriate for our age-specific analyses. The novel findings for malaria exposure and CPRD-recorded EBV infection warrant further study using better methods, such as detailed questionnaires, serology or record linkage. Finally, despite the matched design, the cases and controls had somewhat different average years of observation in the CPRD, which could introduce biases; the fact that the findings did not change with adjustment for years of observation was reassuring.

Conclusion

Our exploratory study of BL aetiology in the UK using the CPRD population identified risk factors that were unexpected, either due to their more conventional association with eBL or immunosuppression-related BL (e.g., HIV/AIDS, EBV infection, apparent exposure to malaria in persons aged 20–59 years) or to the uncertain mechanism that would explain an inverse (cigarette smoking in adults of any age) or positive (prednisone use in persons aged 60 years or older) association. If confirmed, the infection-related findings suggest that patients with poorly controlled EBV infection, a history of infectious mononucleosis, and/or a history of exposure to malaria have an elevated risk of BL, although the absolute risk is probably low.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Ms. Emily Carver, Ruth Parsons, and David Ruggieri at Information Management Services Inc. (Calverton, MD, USA) for coordinating data and preparing data analysis files. This work supported by contracts from the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI) (contract number: HHSN261201500328P) National Institutes of Health, Department of Health and Human Services. The study is based in part on data from the Clinical Practice Research Datalink GOLD database obtained from the UK Medicines and Healthcare Products Regulatory Agency. The content of this publication, the interpretation, and the conclusions contained in this study are those of the author/s alone and do not reflect the views of the UK Medicines and Healthcare Products Regulatory Agency or the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government. The funders did not influence on the study design, the analysis, the interpretation of data, the writing of the report or the decision to submit the article for publication.

References

- Aka P, Vila MC, Jariwala A, Nkrumah F, Emmanuel B, Yagi M, Palacpac NM, Periago MV, Neequaye J, Kiruthu C, Tougan T, Levine PH, Biggar RJ, Pfeiffer RM, Bhatia K, Horii T, Bethony JM, Mbulaiteye SM. Endemic Burkitt lymphoma is associated with strength and diversity of *Plasmodium falciparum* malaria stage-specific antigen antibody response. *Blood*. 2013; 122:629–635. [PubMed: 23645841]
- Anderson LA, Lauria C, Romano N, Brown EE, Whitby D, Graubard BI, Li Y, Messina A, Gafa L, Vitale F, Goedert JJ. Risk factors for classical Kaposi sarcoma in a population-based case-control study in Sicily. *Cancer Epidemiology, Biomarkers & Prevention*. 2008; 17:3435–3443.
- Booth HP, Prevost AT, Gulliford MC. Validity of smoking prevalence estimates from primary care electronic health records compared with national population survey data for England, 2007 to 2011. *Pharmacoepidemiology and Drug Safety*. 2013; 22:1357–1361. [PubMed: 24243711]
- Carpenter LM, Newton R, Casabonne D, Ziegler J, Mbulaiteye S, Mbidde E, Wabinga H, Jaffe H, Beral V. Antibodies against malaria and Epstein-Barr virus in childhood Burkitt lymphoma: a case-control study in Uganda. *International Journal of Cancer. Journal International Du Cancer*. 2008; 122:1319–1323. [PubMed: 18000823]
- Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, Moore PS. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science*. 1994; 266:1865–1869. [PubMed: 7997879]
- Coutinho AE, Chapman KE. The anti-inflammatory and immunosuppressive effects of glucocorticoids, recent developments and mechanistic insights. *Molecular Cell Endocrinology*. 2011; 335:2–13.
- de-The G, Geser A, Day NE, Tukei PM, Williams EH, Beri DP, Smith PG, Dean AG, Bronkamm GW, Feorino P, Henle W. Epidemiological evidence for causal relationship between Epstein-Barr virus and Burkitt's lymphoma from Ugandan prospective study. *Nature*. 1978; 274:756–761. [PubMed: 210392]
- Dregan A, Moller H, Murray-Thomas T, Gulliford MC. Validity of cancer diagnosis in a primary care database compared with linked cancer registrations in England. Population-based cohort study. *Cancer Epidemiology*. 2012; 36:425–429. [PubMed: 22727737]
- Gares V, Panico L, Castagne R, Delpierre C, Kelly-Irving M. The role of the early social environment on Epstein Barr virus infection: a prospective observational design using the Millennium Cohort Study. *Epidemiology and Infection*. 2017:1–8.
- Goedert JJ, Vitale F, Lauria C, Serraino D, Tamburini M, Montella M, Messina A, Brown EE, Rezza G, Gafa L, Romano N. Classical Kaposi's Sarcoma Working G. Risk factors for classical Kaposi's sarcoma. *Journal of the National Cancer Institute*. 2002; 94:1712–1718. [PubMed: 12441327]

- Guech-Ongey M, Simard EP, Anderson WF, Engels EA, Bhatia K, Devesa SS, Mbulaiteye SM. AIDS-related Burkitt lymphoma in the United States: what do age and CD4 lymphocyte patterns tell us about etiology and/or biology? *Blood*. 2010; 116:5600–5604. [PubMed: 20813897]
- Havelange V, Pepermans X, Ameys G, Theate I, Callet-Bauchu E, Barin C, Penther D, Lippert E, Michaux L, Mugneret F, Dastugue N, Raphael M, Vikkula M, Poirel HA. Genetic differences between paediatric and adult Burkitt lymphomas. *British Journal of Haematology*. 2016; 173:137–144. [PubMed: 26887776]
- Hernan MA, Clayton D, Keiding N. The Simpson's paradox unraveled. *International Journal of Epidemiology*. 2011; 40:780–785. [PubMed: 21454324]
- Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, Smeeth L. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *International Journal of Epidemiology*. 2015; 44:827–836. [PubMed: 26050254]
- Hjalgrim H, Askling J, Rostgaard K, Hamilton-Dutoit S, Frisch M, Zhang JS, Madsen M, Rosdahl N, Konradsen HB, Storm HH, Melbye M. Characteristics of Hodgkin's lymphoma after infectious mononucleosis. *New England Journal of Medicine*. 2003; 349:1324–1332. [PubMed: 14523140]
- Klein G, Pearson G, Henle G, Henle W, Diehl V, Niederman JC. Relation between Epstein-Barr viral and cell membrane immunofluorescence in Burkitt tumor cells. II. Comparison of cells and sera from patients with Burkitt's lymphoma and infectious mononucleosis. *The Journal of Experimental Medicine*. 1968; 128:1021–1030. [PubMed: 4878907]
- Kuo CF, Grainge MJ, Mallen C, Zhang W, Doherty M. Impact of gout on the risk of atrial fibrillation. *Rheumatology (Oxford)*. 2016; 55:721–728. [PubMed: 26683193]
- Legason ID, Pfeiffer RM, Udquim KI, Bergen AW, Gouveia MH, Kirimunda S, Otim I, Karlins E, Kerchan P, Nabalende H, Bayanjargal A, Emmanuel B, Kagwa P, Talisuna AO, Bhatia K, Yeager M, Biggar RJ, Ayers LW, Reynolds SJ, Goedert JJ, Ogwang MD, Fraumeni JF Jr, Prokunina-Olsson L, Mbulaiteye SM. Evaluating the Causal Link Between Malaria Infection and Endemic Burkitt Lymphoma in Northern Uganda: A Mendelian Randomization Study. *EBioMedicine*. 2017; 25:58–65. [PubMed: 29033373]
- Leoncini, L.Raphael, M.Stein, H.Harris, N.L.Jaffe, ES., Kluin, PM., editors. *Burkitt lymphoma*. International Agency for Research on Cancer (IARC); Lyon: 2008.
- Levine PH, Kamaraju LS, Connelly RR, Berard CW, Dorfman RF, Magrath I, Easton JM. The American Burkitt's Lymphoma Registry: eight years' experience. *Cancer*. 1982; 49:1016–1022. [PubMed: 7059918]
- Li FY, Chaigne-Delalande B, Su H, Uzel G, Matthews H, Lenardo MJ. XMEN disease: a new primary immunodeficiency affecting Mg²⁺ regulation of immunity against Epstein-Barr virus. *Blood*. 2014; 123:2148–2152. [PubMed: 24550228]
- Mace KE, Arguin PM. Malaria Surveillance - United States, 2014. *MMWR Surveill Summ*. 2017; 66:1–24.
- Macmahon B. Epidemiological evidence of the nature of Hodgkin's disease. *Cancer*. 1957; 10:1045–1054. [PubMed: 13472655]
- Mbulaiteye SM, Anderson WF. Age-related heterogeneity of Burkitt lymphoma. *British Journal of Haematology*. 2018; 180:153–155. [PubMed: 27469517]
- Mbulaiteye SM, Biggar RJ, Goedert JJ, Engels EA. Immune deficiency and risk for malignancy among persons with AIDS. *Journal of the Acquired Immune Deficiency Syndrome*. 2003; 32:527–533.
- Mbulaiteye SM, Anderson WF, Bhatia K, Rosenberg PS, Linet MS, Devesa SS. Trimodal age-specific incidence patterns for Burkitt lymphoma in the United States, 1973–2005. *International Journal of Cancer. Journal International Du Cancer*. 2010; 126:1732–1739. [PubMed: 19810101]
- Mbulaiteye SM, Anderson WF, Ferlay J, Bhatia K, Chang C, Rosenberg PS, Devesa SS, Parkin DM. Pediatric, elderly, and emerging adult-onset peaks in Burkitt's lymphoma incidence diagnosed in four continents, excluding Africa. *American Journal of Hematology*. 2012; 87:573–578. [PubMed: 22488262]
- Mbulaiteye SM, Clarke CA, Morton LM, Gibson TM, Pawlish K, Weisenburger DD, Lynch CF, Goodman MT, Engels EA. Burkitt lymphoma risk in U.S. solid organ transplant recipients. *American Journal of Hematology*. 2013; 88:245–250. [PubMed: 23386365]

- Mbulaiteye SM, Morton LM, Sampson JN, Chang ET, Costas L, de Sanjose S, Lightfoot T, Kelly J, Friedberg JW, Cozen W, Marcos-Gragera R, Slager SL, Birmann BM, Weisenburger DD. Medical history, lifestyle, family history, and occupational risk factors for sporadic Burkitt lymphoma/leukemia: the Interlymph Non-Hodgkin Lymphoma Subtypes Project. *Journal of the National Cancer Institute. Monographs.* 2014a; 2014:106–114. [PubMed: 25174031]
- Mbulaiteye SM, Pullarkat ST, Nathwani BN, Weiss LM, Rao N, Emmanuel B, Lynch CF, Hernandez B, Neppalli V, Hawes D, Cockburn MG, Kim A, Williams M, Altekruze S, Bhatia K, Goodman MT, Cozen W. Epstein-Barr virus patterns in US Burkitt lymphoma tumors from the SEER residual tissue repository during 1979–2009. *APMIS: Acta Pathologica, Microbiologica, Et Immunologica Scandinavica.* 2014b; 122:5–15.
- Nawar E, Mbulaiteye SM, Gallant JE, Wohl DA, Ardini M, Hendershot T, Goedert JJ, Rabkin CS. Risk factors for Kaposi's sarcoma among HHV-8 seropositive homosexual men with AIDS. *International Journal of Cancer. Journal International Du Cancer.* 2005; 115:296–300. [PubMed: 15688390]
- Ogwang MD, Bhatia K, Biggar RJ, Mbulaiteye SM. Incidence and geographic distribution of endemic Burkitt lymphoma in northern Uganda revisited. *International Journal of Cancer. Journal International Du Cancer.* 2008; 123:2658–2663. [PubMed: 18767045]
- Parkin DM, Sitas F, Chirenje M, Stein L, Abratt R, Wabinga H. Part I: Cancer in Indigenous Africans--burden, distribution, and trends. *The Lancet Oncology.* 2008; 9:683–692. [PubMed: 18598933]
- Rochford R, Cannon MJ, Moormann AM. Endemic Burkitt's lymphoma: a polymicrobial disease? *Nature Review Microbiology.* 2005; 3:182–187. [PubMed: 15685227]
- Schlagenhauf P, Weld L, Goorhuis A, Gautret P, Weber R, von Sonnenburg F, Lopez-Velez R, Jensenius M, Cramer JP, Field VK, Odolini S, Gkrania-Klotsas E, Chappuis F, Malvy D, van Genderen PJ, Mockenhaupt F, Jaureguierry S, Smith C, Beeching NJ, Ursing J, Rapp C, Parola P, Grobusch MP. for the EuroTravNet Network. Travel-associated infection presenting in Europe (2008–12): an analysis of EuroTravNet longitudinal, surveillance data, and evaluation of the effect of the pre-travel consultation. *The Lancet Infectious Diseases.* 2015; 15:55–64. [PubMed: 25477022]
- Tatem AJ, Jia P, Ordanovich D, Falkner M, Huang Z, Howes R, Hay SI, Gething PW, Smith DL. The geography of imported malaria to non-endemic countries: a meta-analysis of nationally reported statistics. *The Lancet Infectious Diseases.* 2017; 17:98–107. [PubMed: 27777030]
- Wang Y, Pfeiffer RM, Alsaggaf R, Meeraus W, Gage JC, Anderson LA, Bremer RC, Nikolenko N, Lochmuller H, Greene MH, Gadalla SM. Risk of skin cancer among patients with myotonic dystrophy type 1 based on primary care physician data from the U.K. *Clinical Practice Research Datalink. International Journal of Cancer.* 2018; 142:1174–1181. [PubMed: 29114849]
- Warne B, Weld LH, Cramer JP, Field VK, Grobusch MP, Caumes E, Jensenius M, Gautret P, Schlagenhauf P, Castelli F, Laloo DG, Ursing J, Chappuis F, von Sonnenburg F, Lopez-Velez R, Rapp C, Smith KC, Parola P, Gkrania-Klotsas E. for the EuroTravNet Network. Travel-related infection in European travelers, EuroTravNet 2011. *Journal Of Travel Medicine.* 2014; 21:248–254. [PubMed: 24750378]

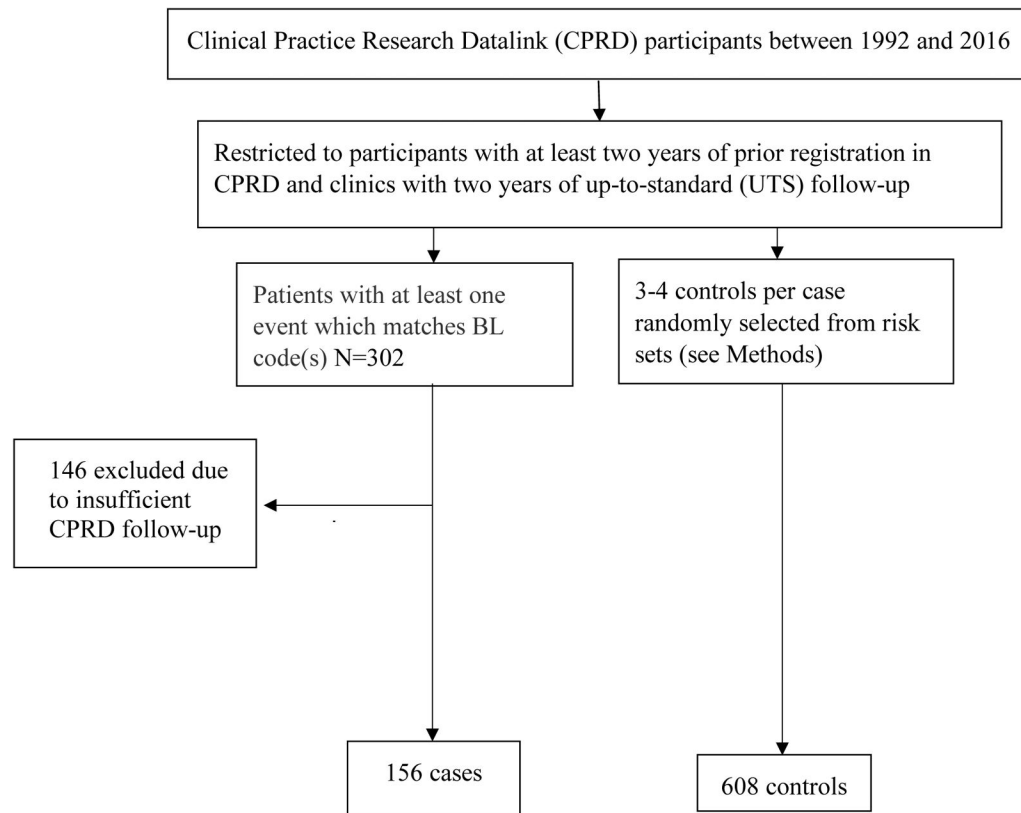


Figure 1.
Flow chart showing the selection of Burkitt lymphoma (BL) cases and controls from the Clinical Practice Research Datalink
The BL cases comprised all patients with a “Read code” for a BL diagnosis and who met CPRD-related eligibility criteria (see Methods)

Table I

Demographic and lifestyle characteristics of the CPRD study sample.

	Cases (N=156)	Controls (N=608)	OR (95%CI)	P
Sex ^a				0.86
Male	112 (71.8)	432 (71.0)		
Female	44 (28.2)	176 (29.0)		
Age at index/selection date, years ^a				0.92
Mean (SD)	48.4 (23.3)	48.2 (23.2)		
Range	2.0–87.0	2.0–87.0		
Age group at index date, years ^a				0.99
<20	27 (17.3)	108 (17.8)		
20–59	69 (44.2)	268 (44.1)		
60+	60 (38.5)	232 (38.2)		
Year at index date ^a				0.99
1992–1999	14 (9.0)	52 (8.5)		
2000–2005	29 (18.6)	112 (18.4)		
2006–2010	58 (37.2)	228 (37.5)		
2011–2016	55 (35.3)	216 (35.5)		
Years since registered in the CPRD ^b				0.005
mean (SD)	17.4 (15.3)	20.9 (13.9)		
Region of residence ^c				0.99
West Midlands	20 (12.8)	80 (13.2)		
South Central	21 (13.5)	76 (12.5)		
Scotland	20 (12.8)	80 (13.2)		
Others	95 (60.9)	372 (61.1)		
BMI (Tertiles)				
T1 (≤ 25.08)	33 (21.2)	97 (16.0)	Ref.	
T2 (25.08 to < 28.82)	26 (16.7)	102 (16.8)	0.73 (0.40 – 1.34)	0.31
T3 (> 28.82)	29 (18.6)	98 (16.1)	0.91 (0.51 – 1.61)	0.74
(Missing: 379)	68 (43.6)	311 (51.1)	N/A	
P _{trend}				0.99
Smoking status before diagnosis				
Never	50 (32.0)	106 (17.4)	Ref.	
Former	64 (41.0)	246 (40.5)	0.47 (0.29 – 0.76)	0.002
Current	15 (9.6)	121 (19.9)	0.24 (0.12 – 0.46)	<0.0001
(Missing: 162) ^d	27 (17.3)	135 (22.2)	N/A	
P _{trend}				<0.0001

	Cases (N=156)	Controls (N=608)	OR (95%CI)	P
Alcohol status before diagnosis				
Former/current	15 (9.6)	38 (6.2)	Ref.	
Never	3 (1.9)	27 (6.6)	1.43 (0.75 – 2.73)	0.274
(Missing: 238)	40 (25.7)	198 (32.6)	N/A	
P_{trend}				

Controls were matched to cases based on age at cohort entry (defined as two years before index/selection date), gender, race, general practice, and index date.

^aDue to matching, odds ratios, and 95% confidence intervals not assessed.

^bDuration of registration in CPRD before the index date. Characteristics were evaluated at least two years before the index date.

^cAbout 40% of participants were from three regions (contributing most cases), including West Midlands, South Central, & Scotland.

^dMajority of the patients missing smoking data were aged <20 years old.

Abbreviations: 95% CI: 95% Confidence Interval; BMI: body mass index; CPRD: Clinical Practice Research Datalink; N/A: not applicable: due to missing data, odds ratios and 95% confidence intervals not assessed; OR: Odds ratio; P: P value for heterogeneity; P_{trend} : P value for trend across categories; Ref.: Reference category; SD: standard deviation; T1, T2, T3: First, second and third tertile, based on cut off values calculated in the controls

Table II

Association of infections and other study selected medical conditions with BL risk.

	Cases (N=156)	Controls (N=608)	OR (95% CI)	P
Any Infection				
Never	22 (14.1)	110 (18.1)	Ref.	
Ever	134 (85.9)	498 (81.9)	1.37 (0.80 – 2.35)	0.25
Hepatitis				
Never	137 (87.8)	562 (92.4)	Ref.	
Ever	19 (12.2)	46 (7.6)	1.88 (1.02 – 3.47)	0.04
EBV				
Never	147 (94.2)	600 (98.7)	Ref.	
Ever	9 (5.8)	8 (1.3)	4.50 (1.74 – 11.7)	0.002
HIV/AIDS				
Never	137 (87.9)	604 (99.3)	Ref.	
Ever	19 (12.2)	4 (0.7)	24.5 (7.25 – 83.0)	<0.0001
Malaria				
Never	151 (96.8)	598 (98.4)	Ref.	
Ever	5 (3.2)	10 (1.6)	2.00 (0.68 – 5.85)	0.21
Immune-related conditions				
Allergy				
Never	120 (76.9)	489 (80.4)	Ref.	
Ever	36 (23.1)	119 (19.6)	1.32 (0.84 – 2.07)	0.23
Hay Fever				
Never	145 (93.0)	565 (93.0)	Ref.	
Ever	11 (7.0)	43 (7.0)	1.02 (0.51 – 2.05)	0.94
Asthma				
Never	132 (84.6)	523 (86.0)	Ref.	
Ever	24 (15.4)	85 (14.0)	1.16 (0.70 – 1.94)	0.56
Eczema				
Never	128 (82.0)	469 (77.1)	Ref.	
Ever	28 (18.0)	139 (22.9)	0.67 (0.41 – 1.09)	0.11
Gout				
Never	149 (95.5)	587 (97.0)	Ref.	
Ever	7 (4.5)	21 (3.0)	1.39 (0.55 – 3.47)	0.49
Use of Prednisone				
Never	147 (94.2)	595 (97.9)	Ref.	
Ever	9 (5.8)	13 (2.1)	3.38 (1.30 – 8.75)	0.01

Characteristics were evaluated at least two years before the index/selection date and classified (ever/never) according to the presence or absence of a “Read code” consistent with a history of the given characteristic. ‘Any infections’ was created based on having a Read code for specified infections due to viruses, bacteria, parasites, or infectious syndromes (See Methods). Thus, it is not the sum of the listed infections.

Abbreviations: 95% CI: 95% Confidence Interval; AIDS: acquired immune deficiency syndrome; BL: Burkitt lymphoma; EBV: Epstein–Barr virus; HIV: human immunodeficiency virus; OR: Odds ratio; P: P value for heterogeneity; Ref.: Reference category.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table III

Association of putative risk factors with BL risk in participants aged 20–59 years and 60+ years.

	20–59 years				60+ years							
	Case (N=69)	Control (N=268)	OR (95% CI)	P	aOR (95% CI)*	P	Case (N=60)	Control (N=232)	OR (95%CI)	P	aOR (95% CI)**	P
BMI[†]												
T1	14	40	Ref.				17	55	Ref.			
T2	12	42	0.80 (0.32 – 2.04)	0.65			12	58	0.64 (0.27 – 1.53)	0.31		
T3	14	40	1.07 (0.44 – 2.56)	0.89			13	55	0.80 (0.35 – 1.84)	0.60		
Missing	29	146	N/A				18	64	N/A			
Trend				0.63						0.73		
Smoking												
Never	32	54	Ref.		Ref.		15	41	Ref.		Ref.	
Former	22	111	0.25 (0.12 – 0.50)	0.001	0.22 (0.09 – 0.52)	<0.001	41	134	0.76 (0.36 – 1.58)	0.61	0.61 (0.27 – 1.34)	0.22
Current	11	74	0.20 (0.09 – 0.46)	<0.001	0.12 (0.04 – 0.35)	<0.001	3	45	0.18 (0.05 – 0.67)	0.01	0.14 (0.03 – 0.58)	0.007
Missing	4	29	N/A		N/A	N/A	1	12	N/A		N/A	
P _{trend}				<.0001		<0.001				0.013		0.007
Any Infection												
Never	13	76	Ref.				7	27	Ref.			
Ever	56	192	1.76 (0.87 – 3.54)	0.12			53	205	0.95 (0.37 – 2.47)	0.92		
Hepatitis												
Never	56	232	Ref.				54	223	Ref.		Ref.	
Ever	13	36	1.59 (0.77 – 3.31)	0.21			6	9	3.41 (1.01 – 11.5)	0.05	3.21 (0.89 – 11.6)	0.08
EBV												
Never	62	263	Ref.		Ref.		58	230	Ref.		Ref.	
Ever*	7	5	5.60 (1.78 – 17.6)	0.003	5.69 (1.29 – 25.1)	0.021	2	2	4.00 (0.56 – 28.4)	0.17	11.8 (0.96 – 144.2)	0.05

	20–59 years				60+ years							
	Case (N=69)	Control (N=268)	OR (95% CI)	P	aOR (95% CI)*	P	Case (N=60)	Control (N=232)	OR (95% CI)	P	aOR (95% CI)**	P
HIV/AIDS												
Never	53	264	Ref.	Ref.	Ref.	Ref.	57	232				
Ever*	16	4	20.6 (5.98 – 70.7)	<0.001	40.4 (7.74 – 211.1)	<0.0001	3	0				
Malaria												
Never	65	266	Ref.	Ref.	Ref.	Ref.	59	225	Ref.	Ref.		
Ever	4	2	8.00 (1.46 – 43.7)	0.016	44.0 (2.63 – 735.3)	0.008	NS	NS	0.57 (0.07 – 4.64)	0.60		
Use of Prednisone												
Never	67	262	Ref.	Ref.	Ref.	Ref.	53	225	Ref.	Ref.		
Ever	2	6	1.33 (0.27 – 6.60)	0.73			7	7	7.30 (1.81 – 29.4)	0.009	7.78 (1.85 – 32.5)	0.005

* The estimates are from multivariable models including smoking, EBV, HIV/AIDS, malaria controlling for one another, and adjusted for years of observation in the CPRD.

** The estimates are from multivariable models including smoking, hepatitis, EBV, use of prednisone controlling for one another, and adjusted for years of observation in the CPRD.

† BMI tertile values for adults aged 20–59 years are: T1 (<24.63), T2 (>=24.63 to 28.88) and T3 (>28.88); the corresponding values for adults aged 60+ years are T1 (<25.57), T2 (>=25.57 to 28.74) and T3 (>=28.74)

Abbreviations: 95% CI: 95% Confidence Interval; AIDS: acquired immune deficiency syndrome; BL: Burkitt lymphoma; BMI: body mass index; CPRD: Clinical Practice Research Datalink; EBV: Epstein-Barr virus; HIV: human immunodeficiency virus; N/A: not applicable due to missing data; NS: cell numbers suppressed because one cell has one subject; OR: Odds ratio; P: P value for heterogeneity; P_{trend}: P value for trend across categories; Ref.: Reference category