

---

## Associations between infectious mononucleosis and cancer: record-linkage studies

---

M. J. GOLDACRE\*, C. J. WOTTON AND D. G. R. YEATES

*Unit of Health-Care Epidemiology, Department of Public Health, University of Oxford, Oxford, UK*

*(Accepted 30 July 2008; first published online 8 October 2008)*

### SUMMARY

Infection with Epstein–Barr virus (EBV) followed by infectious mononucleosis (IM) is now considered to be a risk factor for Hodgkin’s disease (HD). It is less clear whether EBV infection and IM are associated with an increased risk of cancer generally. We used a longstanding record-linkage dataset in Oxford (years 1963–1998), and a more recent record-linkage dataset covering England (1999–2005), to compare rate ratios for cancer between people admitted to hospital for IM and a reference cohort. In the Oxford cohort, there was an increased risk of subsequent HD [rate ratio (RR) 6·0, 95% confidence interval (CI) 2·4–12·5] but not of other cancers combined (RR 0·85, 95% CI 0·57–1·23). In the England cohort, there were increased risks of HD (RR 3·2, 95% CI 1·2–7·0), non-Hodgkin’s lymphoma (RR 5·6, 95% CI 2·9–9·8), and oropharyngeal cancer (RR 5·4, 95% CI 1·1–16·2), but no significant overall risk of cancer when lymphomas were excluded (RR 1·01, 95% CI 0·71–1·41). We confirm an association between IM and lymphoma; but the risk, if any, of cancer more generally is likely to be small.

**Key words:** Cancer, Epstein–Barr virus, infectious mononucleosis, lymphoma, record linkage.

### INTRODUCTION

Infectious mononucleosis (IM) is a clinical syndrome caused by infection with Epstein–Barr virus (EBV). In Western countries a substantial percentage of people have been infected by EBV at some stage in their lives, mostly in infancy, but few develop IM. IM typically develops following relatively late EBV infection, after infancy and early childhood, in adolescence and early adult life. Infection with EBV, and the development of IM, is now considered to be a risk factor for Hodgkin’s disease (HD) [1–4]. Infection with EBV is

also associated with an increased risk of Burkitt’s lymphoma in Africa and Papua New Guinea, and with nasopharyngeal cancer in Africa and southern China. The question of whether EBV infection is associated with an increased risk of other cancers is less well investigated. It is commonly held that ‘EBV is also found in a number of other malignancies, both lymphoid and non-lymphoid, which suggests that EBV infection may not be specific to Hodgkin’s disease’ [1]. In a recent review, Young & Rickinson, in commenting on EBV and Burkitt’s lymphoma, wrote that, ‘EBV has moved from being a bit-part player in the story of an obscure African tumour to its present leading role as the prime example of a human tumour virus that is aetiologically linked to an unexpectedly diverse range of malignancies’ [5]. Our aim was to

\* Author for correspondence: Professor M. J. Goldacre, Unit of Health-Care Epidemiology, Department of Public Health, University of Oxford, Old Road Campus, Oxford OX3 7LF, UK.  
(Email: michael.goldacre@dphpc.ox.ac.uk)

determine whether there is epidemiological evidence to support the hypothesis that the syndrome of IM is not only associated with lymphoma, but also with a more general elevation of risk of cancer in England. We used two record-linkage datasets – the Oxford Record Linkage Study (ORLS) from 1963–1998 and a dataset for the whole of England from 1999–2005. The rationale for using both is that the ORLS provides information about longer mean periods of observation between IM and cancer than the dataset for England, whereas the dataset for England covers a much larger population but with much shorter follow-up.

## METHODS

### Population and data

We used data from the Oxford Record Linkage Study (ORLS) and from English national Hospital Episode Statistics (HES) [6, 7]. The English NHS Central Office for Research Ethics Committees approved the current work programme of analysis using the datasets (reference number 04/Q2006/176). The ORLS includes brief statistical abstracts of records of all hospital admissions (including day cases) in National Health Service (NHS) hospitals, and all deaths, whether in hospital or not, provided that the death occurred within the geographical area covered by the ORLS. Data collection was undertaken within the former Oxford NHS Region and it covered one health district from 1963, two from 1965 (population 850 000), six from 1975 (population 1.9 million) and eight from 1987 (population 2.5 million), until 1998. Linkable data have been collected since then in the Oxford region, as part of national English linkage, but the linkage keys are not compatible across the 1998 divide. The ORLS hospital data were collected routinely in the NHS as the region's hospital statistics system and were similar to English national HES. The death data derive from death certificates. The data for each individual were linked together by probabilistic linkage using encrypted names, dates of birth, and addresses, as they accrued, and are now fully anonymized and archived. The English data, an extract from linked national HES and death data, spanned 1999–2005. It was linked using encrypted NHS numbers, dates of birth and postcodes.

Using the ORLS, a cohort was constructed of people admitted to hospital, with an International Classification of Diseases (ICD) code for infectious

mononucleosis (IM) on the discharge record, aged <65 years. This was done by identifying the first admission for IM in a NHS hospital in the former Oxford NHS Region during the study period (we termed this the 'exposure cohort' for this analysis). The ICD codes used for IM were 093 in ICD-7 (from 1963 to 1967), 075 in ICD-8 (from 1968 to 1978) and ICD-9 (from 1979 to 1994), and B27 in ICD-10 (from 1994). The main terms covered by the codes are 'glandular fever' and 'infectious mononucleosis'. No information is available in the coding about whether the diagnosis was a clinical one or whether there was diagnostic laboratory confirmation. Terminology was similar over time and any confounding that might have been introduced by differences in diagnostic criteria over time was reduced by the fact that the analyses were stratified and standardized by year of admission (see below).

A reference cohort was constructed by identifying the first admission for each individual with various common and reasonably minor reasons for hospital admission (mainly surgical or orthopaedic conditions, or injuries, see note to Table 1) as the main diagnosis or operation. This is based on a 'reference' group of conditions that has been used in other studies of associations between non-malignant diseases and subsequent cancer [6, 8–10]. We used the standard epidemiological approach of using a variety of conditions in choosing a hospital comparison group (to avoid the possibility of atypical cancer risk in any one). The choice of conditions was made for this and other studies [6, 8–10] by the investigators. We then tested each individual reference condition, by analysis of its own cancer risk, to determine whether any individual condition gave atypically high or low cancer rates. This led us to exclude one common condition in the original group – admissions for upper respiratory tract infection – because they showed some elevated risks of associated cancers.

People were excluded from the IM and reference cohort if they had an admission for cancer either before or at the same time as the admission for IM or the reference condition. We searched the dataset for any subsequent NHS hospital care for, or death from, cancer in these cohorts. We considered that rates of cancer in the reference cohort would approximate those in the general population of the region while allowing for migration in and out of it (data on migration of individuals were not available).

The same methods and selection criteria were used to construct cohorts using the English national data.

Table 1. Number of people admitted to hospital with infectious mononucleosis or with a reference cohort condition in each age-group stratum: data for Oxford 1963–1998 and England 1999–2005

| Age groups<br>(years) | Oxford, 1963–1998           |       |                   |                    | England, 1999–2005          |       |                   |                    |
|-----------------------|-----------------------------|-------|-------------------|--------------------|-----------------------------|-------|-------------------|--------------------|
|                       | Infectious<br>mononucleosis |       | Reference cohort* |                    | Infectious<br>mononucleosis |       | Reference cohort* |                    |
|                       | No.                         | %     | No.               | Matching<br>ratio† | No.                         | %     | No.               | Matching<br>ratio† |
| 0–4                   | 128                         | 4.6   | 45867             | 358                | 654                         | 4.4   | 219036            | 335                |
| 5–9                   | 221                         | 7.9   | 51467             | 233                | 924                         | 6.1   | 350779            | 380                |
| 10–14                 | 325                         | 11.6  | 38792             | 119                | 1798                        | 11.9  | 273238            | 152                |
| 15–19                 | 1032                        | 37.0  | 44846             | 43                 | 6807                        | 45.3  | 225725            | 33                 |
| 20–24                 | 574                         | 20.5  | 53701             | 94                 | 2407                        | 16.0  | 254965            | 106                |
| 25–29                 | 222                         | 7.9   | 50103             | 226                | 948                         | 6.3   | 277224            | 292                |
| 30–34                 | 99                          | 3.5   | 45514             | 460                | 568                         | 3.8   | 293178            | 516                |
| 35–39                 | 67                          | 2.4   | 42344             | 632                | 373                         | 2.5   | 289635            | 777                |
| ≥40                   | 129                         | 4.6   | 170890            | 1325               | 550                         | 3.7   | 1221916           | 2222               |
| Total                 | 2797                        | 100.0 | 543524            | 194                | 15029                       | 100.0 | 3405696           | 227                |

\* The reference cohorts were constructed by identifying the first hospital admission for each person with an ICD code at its discharge, as the principal diagnosis, for any of the following conditions. Conditions used in reference cohort, with Office of Population, Censuses and Surveys (OPCS) code edition 3 for operations and ICD-9 code for diagnosis (with equivalent codes used for other coding editions): tonsillectomy (230), appendicectomy (441), squint (378), otitis externa, otitis media (380–382), varicose veins (454), deflected septum, nasal polyp (470–471), impacted tooth and other disorders of teeth (520–521), inguinal hernia (550), ingrowing toenail and other diseases of nail (703), sebaceous cyst (706.2), internal derangement of knee (717), bunion (727.1), selected fractures (810–816, 823–826), dislocations, sprains and strains (830–839, 840–848), superficial injury and contusion (910–919, 920–924).

† The number of people in the reference cohort per person in the infectious mononucleosis cohort.

## Statistical methods

We calculated rates of each cancer based on person-years at risk. We took ‘date of entry’ into each cohort as the date of first admission for IM, or reference condition, and ‘date of exit’ for the analysis of each individual cancer as the date of first record of cancer, death, or the end of the data file (1998 for the ORLS, 2005 for England), whichever was the earliest. In comparing the IM cohort with the reference cohort, we first calculated rates for cancer, standardized by age (in 5-year age groups), sex, calendar year of first recorded admission, and either district of residence (ORLS) or region of residence (England), taking the combined IM and reference cohorts as the standard population. We then applied the overall rates to the age structure of the individual cohorts of people with IM or with the reference conditions. We calculated the ratio of the standardized rate of occurrence of cancer in the exposure cohorts relative to that in the reference cohort. The calculation of the rate ratios, and their confidence intervals, were based on the methods described by Breslow & Day [11]. In brief, we

considered the observed counts of outcomes of each cancer to be distributed as Poisson variables. Since the ratio of two such variables follows a binomial distribution, we used this distribution to obtain the confidence intervals on the rate ratios.

We repeated the analyses excluding admissions with an ICD code for cancer that occurred within a year following the admission for IM. We did so to reduce the possibility that the cancer was present at the time of IM and to reduce the possibility that, for lymphoma in particular, the tumour in a febrile patient may have initially been diagnosed as ‘glandular fever’.

In comparing the IM and reference cohorts, the precision of the rate ratio depends on the number of people with each subsequent disease within each cohort. The size of the IM cohort is fixed by the number in the database with the condition. In the reference cohort, we included all the people in the database with the reference conditions in each age group. We used all the people in each stratum in the reference cohort in order to maximize the precision of the rate ratios.

### Risk of IM in people with cancer

We also considered the possibility that IM might be more common in people who already had cancer than in others. To investigate this, we reversed the study design: we constructed a cohort of people with cancer (termed the 'exposure' cohort for this analysis) and sought records of subsequent admission for IM in them, to compare with the reference cohort. In this analysis we stratified the cancer cohort and the reference cohort by date of admission for cancer or reference condition, by year of admission for cancer or reference condition, and by sex and district of residence, and conducted the analyses using the statistical methods as described above.

## RESULTS

In the Oxford data, there were 2797 people in the cohort of people with a diagnostic code for IM at hospital discharge, of whom 61% (1706) were aged <20 years and 89% (2502) were aged <30 years at the time of hospital admission. There were 543 524 people in the reference cohort.

In the English data, there were 15 029 people in the cohort of people with a diagnostic code for IM at hospital discharge, of whom 68% (10 183) were aged <20 years and 90% (13 538) were aged <30 years at the time of hospital admission. In both cohorts, the modal age group at admission for IM was 15–19 years, the second most frequent age group was 20–24 years, and the third was 10–14 years. Admission for IM was substantially more common in males than in females in most age groups; more common in females than males at ages 10–14 years; and numbers of males and females were very similar in the commonest age group of 15–19 years. There were 3 405 696 people in the reference cohort.

Table 1 shows the age distribution of people admitted with IM compared to those admitted for one of the reference conditions in both the Oxford and England populations. It shows that there were high matching ratios, of reference cohort conditions to people in the IM cohort, in every age stratum.

### Cancer after IM: Oxford data

The rate ratio for HD was significantly elevated [6.0, 95% confidence interval (CI) 2.4–12.5, based on seven observed cases, Table 2]. The time intervals from IM to HD were 1 year, 2 years, 4 years (two patients), 5 years and 7 years (two patients). The rate ratio for

non-Hodgkin's lymphoma was not significantly elevated (1.8, 95% CI 0.4–5.2, based on three observed cases, Table 2).

There was no elevation of risk of cancer overall in the IM cohort, compared with the reference cohort (Table 2): the rate ratio for all cancers combined was 1.03 (95% CI 0.7–1.4), and excluding the cases of lymphoma, the rate ratio was 0.8 (95% CI 0.5–1.2). In Table 2, we have shown data for each cancer with two or more observed or expected cases. There were no cases of cancer of the nasopharynx (compared with 0.1 expected cases). There were no cases of leukaemia (compared with an expected value of 1.3). Other cancers that were studied but are not included in the table (because both the observed and expected number was <2) were cancers of the oral cavity, pharynx and lip, larynx, salivary gland, oesophagus, stomach, rectum, liver, pancreas, cervix, uterus, ovary, prostate, testis, kidney, malignant and benign brain, other nervous system, thyroid, bone and multiple myeloma.

### Cancer after IM: England data

There was an elevated risk of HD (3.2, 95% CI 1.2–7.0, based on six cases; Table 3). There was also an elevated risk of non-Hodgkin's lymphoma in this cohort (5.6, 95% CI 2.9–9.8), but the elevated risk was confined to admissions of non-Hodgkin's lymphoma that occurred within 1 year of the admission for IM (Table 3). Of the 12 cases of non-Hodgkin's lymphoma, five were coded with the ICD codes for 'peripheral and cutaneous lymphoma', five with the code for 'diffuse non-Hodgkin's lymphoma' and two with the code for 'other and unspecified types'.

There was a significant elevation of risk of cancer overall in the IM cohort, compared with the reference cohort (Table 3): the rate ratio for all cancers combined was 1.4 (95% CI 1.03–1.8). However, exclusion of the cases of lymphoma showed that there was no excess of the other cancers combined [rate ratio (RR) 1.01, 95% CI 0.7–1.4]. In Table 3, we have shown data where there were two or more observed or expected cases. Cancers that were significantly elevated, in addition to lymphoma, included cancer of the pancreas (RR 7.7, 95% CI 1.6–22.7) and prostate (RR 4.9, 95% CI 1.02–14.5). Cancers of the pancreas and prostate were only significantly elevated when the cancer admission occurred within a year of IM admission [rate ratios excluding first-year cases were

Table 2. *Oxford record-linkage study data, 1963–1998: occurrence of cancer in people with infectious mononucleosis (IM) compared with the reference cohort\**

| Cancer (ICD code†)                    | Number in reference cohort with cancer | Number in IM cohort with cancer | Expected number in IM cohort | Adjusted rate ratio‡ | 95 % CI   |
|---------------------------------------|--|---------------------------------|------------------------------|----------------------|-----------|
| All cancers (140–208)                 | 20451                                  | 35                              | 34.1                         | 1.03                 | 0.72–1.43 |
| All cancers excluding 1st year cases§ | 17871                                  | 31                              | 31.9                         | 0.97                 | 0.66–1.38 |
| All cancers excluding lymphoma        | 19574                                  | 25                              | 31.2                         | 0.80                 | 0.52–1.18 |
| Individual cancers                    |  |                                 |                              |                      |           |
| Rectum (154)                          | 1076                                   | 2                               | 1.44                         | 1.39                 | 0.17–5.04 |
| Colon (153)                           | 1750                                   | 1                               | 2.45                         | 0.41                 | 0.01–2.28 |
| Breast (174) (females only)           | 3369                                   | 6                               | 5.25                         | 1.14                 | 0.42–2.48 |
| Lung (162)                            | 2896                                   | 1                               | 3.28                         | 0.30                 | 0.01–1.70 |
| Testis (186)                          | 162                                    | 2                               | 1.30                         | 1.55                 | 0.19–5.69 |
| Bladder (188)                         | 1167                                   | 3                               | 1.72                         | 1.74                 | 0.36–5.10 |
| All skin (172, 173)                   | 1806                                   | 3                               | 2.86                         | 1.05                 | 0.22–3.07 |
| Hodgkin's disease (201)               | 183                                    | 7                               | 1.21                         | 5.96                 | 2.36–12.5 |
| Non-Hodgkin's lymphoma (200, 202)     | 694                                    | 3                               | 1.69                         | 1.78                 | 0.37–5.23 |

\* See asterisked note in Table 1.

† ICD-9 codes for each cancer (equivalent codes were used for cases coded in ICD revisions 7, 8 and 10).

‡ Ratio of the rate in the IM cohort compared with the rate in the reference cohort, adjusted for sex, age in 5-year bands, time-period in single calendar years, and district of residence.

§ People with an admission with an ICD code for cancer, within 1 year of admission for IM, were excluded from this analysis.

Table 3. *English record-linkage data, 1999–2005: occurrence of cancer in people with infectious mononucleosis (IM) compared with the reference cohort\**

| Cancer (ICD code†)                      | Number in reference cohort with cancer | Number in IM cohort with cancer | Expected number in IM cohort | Adjusted rate ratio‡ | 95 % CI   |
|---|--|---------------------------------|------------------------------|----------------------|-----------|
| All cancers (140–208)                   | 39232                                  | 53                              | 38.7                         | 1.37                 | 1.03–1.79 |
| All cancers excluding first-year cases§ | 26877                                  | 30                              | 25.8                         | 1.16                 | 0.78–1.66 |
| All cancers excluding lymphoma          | 37083                                  | 35                              | 34.5                         | 1.01                 | 0.71–1.41 |
| Individual cancers                      |  |                                 |                              |                      |           |
| Oral cavity, pharynx, lip (140–149)     | 913                                    | 3                               | 0.50                         | 5.54                 | 1.14–16.2 |
| Colon (153)                             | 2739                                   | 3                               | 2.51                         | 1.19                 | 0.25–3.48 |
| Pancreas (157)                          | 728                                    | 3                               | 0.47                         | 7.73                 | 1.59–22.7 |
| Breast (174) (females only)             | 5975                                   | 4                               | 3.01                         | 1.35                 | 0.37–3.45 |
| Prostate (185)                          | 2357                                   | 3                               | 0.60                         | 4.94                 | 1.02–14.5 |
| All skin (172, 173)                     | 5506                                   | 3                               | 4.19                         | 0.72                 | 0.15–2.09 |
| Malignant brain (191)                   | 962                                    | 4                               | 1.91                         | 2.15                 | 0.58–5.52 |
| Hodgkin's disease (201)                 | 467                                    | 6                               | 1.95                         | 3.21                 | 1.17–7.04 |
| Non-Hodgkin's lymphoma (200, 202)       | 1682                                   | 12                              | 2.23                         | 5.59                 | 2.88–9.79 |
| Multiple myeloma (203)                  | 467                                    | 1                               | 0.32                         | 3.99                 | 0.10–22.3 |
| Leukaemia (204–208)                     | 1157                                   | 5                               | 2.21                         | 2.23                 | 0.72–5.23 |

\* See asterisked note in Table 1.

† ICD-9 codes for each cancer (equivalent codes were used for cases coded in ICD revisions 7, 8 and 10).

‡ Ratio of the rate in the IM cohort compared with the rate in the reference cohort, adjusted for sex, age in 5-year bands, time-period in single calendar years, and district of residence.

§ People with an admission with an ICD code for cancer, within 1 year of admission for IM, were excluded from this analysis.

Table 4. Occurrence of cancer, excluding lymphoma, in people with infectious mononucleosis (IM) compared with the reference cohort: time intervals between admission for IM and admission for cancer

| Study location and time intervals | Number in IM cohort with cancer | Expected number in IM cohort | Adjusted rate ratio* | 95 % CI   |
|-----------------------------------|---------------------------------|------------------------------|----------------------|-----------|
| Oxford, 1963–1998                 |                                 |                              |                      |           |
| <1 yr                             | 3                               | 3.0                          | 1.00                 | 0.21–2.92 |
| 1–4 yr                            | 5                               | 4.2                          | 1.19                 | 0.39–2.78 |
| 5–9 yr                            | 5                               | 4.7                          | 1.06                 | 0.35–2.48 |
| ≥10 yr                            | 12                              | 19.3                         | 0.62                 | 0.32–1.09 |
| England, 1999–2005                |                                 |                              |                      |           |
| <1 yr                             | 10                              | 11.4                         | 0.88                 | 0.42–1.61 |
| 1–4 yr                            | 19                              | 16.8                         | 1.13                 | 0.68–1.68 |
| ≥5 yr                             | 6                               | 6.3                          | 0.95                 | 0.35–2.07 |

\* Ratio of the rate in the IM cohort compared with the rate in the reference cohort, adjusted for sex, age in 5-year bands, time-period in single calendar years, and district of residence.

0 (95 % CI 0–15.2) and 2.01 (95 % CI 0.05–11.2), respectively]. There was also an elevated risk of cancer of the oral cavity (RR 5.5, 95 % CI 1.1–16.2; Table 3). The observed cases were two cases coded as cancer of the hypopharynx and one case coded as cancer of the tongue. All three cases of these cancers occurred within 2 years of the admission for IM. There were no cases of cancer of the nasopharynx (compared with 0.2 expected cases). Cancers that were studied but are not included in the table (because both the observed and expected number was <2) were cancers of the larynx, salivary gland, oesophagus, stomach, rectum, liver, lung, cervix, uterus, ovary, testis, kidney, benign brain, other nervous system, thyroid and bone.

#### Time intervals

If IM leads to an increased risk of non-lymphoma cancer, a time lag of several years might be expected (allowing for latency). Accordingly, we analysed time intervals (Table 4). Considering the combined group of cancers excluding lymphoma, there is no evidence that the rate ratio for cancer increases with time from IM admission in either the Oxford or the England population (Table 4). Considering HD, all but one of the Oxford cases occurred >1 year but <8 years after IM. In the England data (with much shorter potential for follow-up than in the Oxford data), all six people with HD developed it within 5 years of IM. Ten of the 12 England cases of non-Hodgkin's lymphoma were first admitted to hospital with it within a year of admission for IM.

#### IM after cancer: Oxford data

In this analysis, we confined the 'exposure cohort' of people with cancer to those whose first admission for cancer was at age <30 years (because almost 90% of all cases of IM were aged <30 years). There were 7038 people with cancer and four people with a subsequent admission for IM, compared with an expected number of 3.9 (RR 1.04, 95 % CI 0.3–2.7). There were no cases of admission for IM after a diagnosis of lymphoma (0.5 expected).

#### IM after cancer: England data

This analysis, too, was confined to those with a first admission for cancer aged <30 years (because 90% of all cases of IM in the England data were aged <30 years). There were 59 905 people with cancer and 21 people with a subsequent admission for IM, compared with an expected number of 15.4 (RR 1.4, 95 % CI 0.9–2.1). The rate ratio for IM after lymphoma was 2.2 (95 % CI 0.7–5.1, based on five observed and 2.3 expected cases).

#### DISCUSSION

Interest in the possibility that virus infection could cause cancer was first awakened by the work of Peyton Rous in the early decades of the 20th century. He was eventually awarded the Nobel Prize in 1966 for his work on virus transmission and chicken-sarcoma. The role of oncoviruses in the aetiology of some cancers is now well established [12].

Our study has similar findings to that of a large record-linkage study performed in Sweden and Denmark [13]. We found, as they did, an elevated rate ratio, with a relatively short period of follow-up, between IM and HD. We both found an increased rate ratio for non-Hodgkin's lymphoma that was no longer present after the first-year cases were excluded, and both studies found no elevated risk of cancer generally after lymphoma was excluded.

### Lymphoma

The first major epidemiological study of IM and HD was published by Rosdahl *et al.* from Denmark in 1974 [2]. In their study, 17/17073 people who had a positive reaction to the Paul-Bunnell test for IM subsequently developed HD, compared with an expected number of six (a threefold elevation of risk with a *P* value, quoted by the authors, of  $<0.0002$ ). Since then, evidence of an association has accumulated through studies that have shown raised antibody titres to EBV associated with HD, and through studies using specific gene probes for EBV, and polymerase chain reaction [1]. Diepstra *et al.* postulated biological mechanisms through which EBV infection may contribute to the development of HD, concerning genetically determined immune responses to challenge with EBV antigens [3].

The record linkage study from Sweden and Denmark found an association between IM and HD, with a standardized incidence ratio of 2.55 (95% CI 1.87–3.40) [13]. A recent population-based case-control study performed in Italy found an elevated risk of HD (age adjusted odds ratio 4.4, 95% CI 1.1–16.6) [14]. A recent case-control study performed in the UK found a significant association between IM and HD (RR 2.43, 95% CI 1.10–5.33), and found that that EBV-positive HD was significantly more likely to be associated with IM than EBV-negative HD [15]. Our rate ratios for HD are broadly similar to the risk in these studies. We cannot completely rule out the possibility that HD may have been misdiagnosed as IM in any of the cases where the time intervals between them were fairly short. In the Danish and Swedish study, 17/46 cases of HD occurred within 5 years of IM, and 30/46 within 10 years of IM [13]. Hjalgrim *et al.* [16], who studied the characteristics of the Danish cohort in detail, reported that the median time from infection to Hodgkin's lymphoma was 4 years. The median times in our study, in the Oxford and England

datasets, were only 4 and 2 years respectively. Nonetheless, misdiagnosis seems an unlikely explanation for the findings. There is now substantial evidence to support the view that the association between EBV infection and HD is causal. As Mueller has written [12], there are consistent serological and molecular patterns of EBV 'fingerprints' associated with HD. The data on time-periods indicate that latent period from IM to HD can be quite short.

In the England data, we found a significant association with non-Hodgkin's lymphoma as well as with HD. However, the elevation of risk in our study was only found when non-Hodgkin's lymphoma was diagnosed within a year of IM, as it was in the Danish-Swedish study [13]. Misclassification of lymphoma, initially as IM, is a possible explanation for this finding. An alternative explanation, for both the Danish-Swedish findings and ours, is that IM infection results in very rapid oncogenic transformation in respect of non-Hodgkin's lymphoma. EBV causes very marked B cell proliferation. A short latent period between infection and lymphoma is plausible. A further explanation is that the association may be the result of unmeasured confounding – that those who are prone to IM are also prone to non-Hodgkin's lymphoma.

### Other cancers

The Swedish and Danish study [13] reported a significant association between IM and skin cancer overall, and also significant associations, separately, for malignant melanoma and for non-melanoma skin cancer. Numbers of people with skin cancer in our study are too small to comment on this association. In the England data (the larger of the two cohorts, but with limited follow-up), we found significant elevations of cancers of the oral cavity, pancreas and prostate. Neither cancers of the pancreas or prostate have, to our knowledge, been reported before and were only significant in the short-term after IM, suggesting that the association is non-causal. Furthermore, we made multiple comparisons between IM and cancers and a few significant findings would be expected by chance. Our finding of an elevated risk of oropharyngeal cancer supports a previous study on EBV and an elevated risk of oropharyngeal cancer [17], but the large Swedish study did not find an association with oropharyngeal cancer [13].

### Total risk of cancer after IM, excluding lymphoma

Our data, and that from the Danish and Swedish cohort [13], provide strong evidence that there is no general increase in risk of cancer following IM infection.

### Risk of IM in people with cancer

We also studied the risk of hospital admission for IM after an initial diagnosis of cancer. We did so, first, because we considered the possibility that a cancer, perhaps particularly a lymphoma, might initially be diagnosed as IM. Second, we considered the possibility that if cancer and IM are associated, the association might be non-causal but confounded by a shared increase in individual susceptibility to both the cancer and IM. In this case, any association might be expected to be found bi-directionally – an increased risk of IM both before and after cancer. Our findings do not support the hypothesis that there is an elevation of risk of IM after the occurrence of cancer.

### Strengths and weaknesses of the study

A strength of our study is that, using the datasets, we were able to analyse associations with all cancers as well as with HD. Unlike case-control studies based on recall of infections in the distant past, our study is not susceptible to reporting biases. We used both the Oxford and the England datasets to take advantage of the much longer follow-up of the Oxford dataset and the much larger size of the English dataset.

The study design also has a number of limitations. First, we have no clinical data other than the ICD-coded diagnoses. We have no information about whether there was laboratory confirmation of the diagnosis of IM. A further limitation of our study is that it is confined to patients whose infection was serious enough to warrant hospital care. Infection with EBV is very common; only a small minority of people infected with EBV develop IM; and only a minority of people with IM are admitted to hospital. This means that our study design includes some misclassification: it is a comparison between one cohort, all of whom had IM, and another cohort, many of whom will have had EBV infection and IM without hospital admission. This would reduce our chances of detecting a difference between the cohorts, if one truly exists, so it is noteworthy that the expected positive

association between IM and HD was indeed found. We hope that this may encourage others to use similar methodology in pursuing other hypotheses about infectious diseases and their possible sequelae, if the latter require hospital care, using linked administrative hospital statistics.

The study is too small to detect the risk of nasopharyngeal carcinoma. Nasopharyngeal cancer is very rare in England, but associations between it and EBV infection in African and Chinese populations have been well documented [18, 19].

In summary, we add support to the evidence that the risk of HD is increased after IM. Our data also suggest that this association is probably specific to HD, and perhaps oropharyngeal cancer, and that, at least in this relatively young Western population, there is not a more generalized increase in the risk of cancer.

### ACKNOWLEDGEMENTS

The Oxford Record Linkage Study and the English national linked dataset were built under the direction of Leicester Gill. The English National Co-ordinating Centre for Research Capacity Development funds the Unit of Health Care Epidemiology to undertake research using the linked datasets.

### DECLARATION OF INTEREST

None.

### REFERENCES

1. **Bunch C, Gatter K.** Disorders of the blood: the lymphomas. In: Weatherall DJ, Ledingham JGG, Warrell DA, eds. *Oxford Textbook of Medicine*, 3rd edn. Oxford: Oxford University Press, 1996, pp. 3568–3587.
2. **Rosdahl N, Larsen S, Clemmesen J.** Hodgkin's disease in patients with previous infectious mononucleosis: 30 years' experience. *British Medical Journal* 1974; **2**: 253–256.
3. **Diepstra A, et al.** Association with HLA class I in Epstein-Barr-virus-positive and with HLA class III in Epstein-Barr-virus-negative Hodgkin's lymphoma. *Lancet* 2005; **365**: 2216–2224.
4. **Hjalgrim H, et al.** Infectious mononucleosis, childhood social environment and risk of Hodgkin lymphoma. *Cancer Research* 2007; **67**: 2382–2388.
5. **Young LS, Rickinson AB.** Epstein-Barr virus: 40 years on. *Nature Reviews Cancer* 2004; **4**: 757–768.
6. **Goldacre M, et al.** Use of large medical databases to study associations between diseases. *QJM: Monthly*



- Journal of the Association of Physicians* 2000; **93**: 669–675.
7. **Goldacre M, Gill L.** Linkage of hospital episode statistics in England: opportunities for public health. London: Faculty of Public Health ([http://www.fph.org.uk/resources/newsletters/phcom/archive/2005/phcom\\_dec05.pdf](http://www.fph.org.uk/resources/newsletters/phcom/archive/2005/phcom_dec05.pdf)). Accessed 11 March 2008.
  8. **Goldacre MJ, et al.** Cancer following hip and knee arthroplasty: record linkage study. *British Journal of Cancer* 2005; **92**: 1298–1301.
  9. **Goldacre MJ, et al.** Cancer after cholecystectomy: record-linkage cohort study. *British Journal of Cancer* 2005; **92**: 1307–1309.
  10. **Goldacre MJ, et al.** Schizophrenia and cancer: an epidemiological study. *British Journal of Psychiatry* 2005; **187**: 334–338.
  11. **Breslow N, Day N.** *Statistical Methods in Cancer Research, Volume II. The Design and Analysis of Cohort Studies*. Lyon: IARC scientific publication no. 82, 1987, pp. 93–96.
  12. **Mueller N.** Overview: viral agents and cancer. *Environmental Health Perspectives* 1995; **103** (Suppl. 8): 259–261.
  13. **Hjalgrim H, et al.** Risk of Hodgkin's disease and other cancers after infectious mononucleosis. *Journal of the National Cancer Institute* 2000; **92**: 1522–1528.
  14. **Vineis P, et al.** Haematopoietic cancer and medical history: a multicentre case control study. *Journal of Epidemiology and Community Health* 2000; **54**: 431–436.
  15. **Alexander FE, et al.** Risk factors for Hodgkin's disease by Epstein Barr virus (EBV) status: prior infection by EBV and other agents. *British Journal of Cancer* 2000; **82**: 1117–1121.
  16. **Hjalgrim H, et al.** Characteristics of Hodgkin's lymphoma after infectious mononucleosis. *New England Journal of Medicine* 2003; **349**: 1324–1332.
  17. **Szkaradkiewicz A, et al.** Epstein-Barr virus and human papillomavirus infections and oropharyngeal squamous cell carcinomas. *Clinical Experimental Medicine* 2002; **2**: 137–141.
  18. **Doll R, Peto R.** *The Causes of Cancer*. Oxford: Oxford University Press, 1981, pp. 1254–1255.
  19. **Hsu JL, Glaser SL.** Epstein-Barr virus-associated malignancies: epidemiologic patterns and etiologic implications. *Clinical Review Oncology/Hematology* 2000; **34**: 27–53.