Comprehensive evaluation of medical conditions associated with risk of non-Hodgkin lymphoma using Medicare claims ("MedWAS")

Eric A. Engels[1], **Ruth Parsons**[2], **Caroline Besson**[1,3], **Lindsay M. Morton**[1], **Lindsey Enewold**[4], **Winnie Ricker**[2], **Elizabeth L. Yanik**[1], **Hannah Arem**[1], **April A. Austin**[5], and **Ruth M. Pfeiffer**[1]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA

[2]Information Management Services, Inc., Rockville, MD, USA

[3]Université Paris Sud, Faculté de médecine, Le Kremlin-Bicêtre 94270, France

[4]Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD, USA

[5]New York State Cancer Registry, New York State Department of Health, Albany, NY, USA

Abstract

Background—Certain medical conditions affect risk of non-Hodgkin lymphoma (NHL), but the full range of associations is unknown. We implemented a novel method ("medical condition-wide association study," MedWAS) to comprehensively evaluate medical risk factors for NHL documented in administrative health claims.

Methods—Using SEER-Medicare data, we conducted a case-control study comparing NHL cases (N=52,691, age 66+ years, with five subtypes: chronic lymphocytic leukemia/small lymphocytic lymphoma [CLL/SLL], diffuse large B-cell lymphoma [DLBCL], follicular lymphoma, marginal zone lymphoma [MZL], T-cell lymphoma [TCL]) to controls (N=200,000).We systematically screened for associations with 5926 medical conditions documented in Medicare claims more than one year before selection.

Results—Fifty-five conditions were variously associated with NHL. Examples include well-established associations of human immunodeficiency virus, solid organ transplantation, and hepatitis C virus with increased DLBCL risk (odds ratios [ORs] 3.83, 4.27, and 1.74, respectively), and autoimmune conditions with DLBCL and MZL (e.g., ORs of 2.10 and 4.74, respectively, for Sjögren syndrome). Risks for all NHL subtypes were increased after diagnoses of non-melanoma skin cancer (ORs 1.19–1.55), actinic keratosis (1.12–1.25), or hemolytic anemia (1.64–4.07). Nine additional skin conditions increased only TCL risk (ORs 2.20–4.12). Diabetes mellitus was

associated with increased DLBCL risk (OR 1.09). Associations varied significantly across NHL subtypes for 49 conditions (89%).

Conclusion—Using an exploratory method, we found numerous medical conditions associated with NHL risk, and many associations varied across NHL subtypes.

Impact—These results point to etiologic heterogeneity among NHL subtypes. MedWAS is a new method for assessing the etiology of cancer and other diseases.

Keywords

Introduction

Non-Hodgkin lymphoma (NHL) is a common malignancy, with 465,000 incident cases worldwide in 2013 (1). Incidence rises with age, and 57% of US cases occur after age 65 years (2). Although considered a single entity for descriptive purposes, NHL comprises a group of heterogeneous subtypes with distinct clinical presentations and, as is increasingly recognized, differing causal pathways (i.e., etiologic heterogeneity) (3). Common NHL subtypes include tumors derived from B-cells such as diffuse large B-cell lymphoma (DLBCL), chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL), follicular lymphoma (FL), and marginal zone lymphoma (MZL). T-cell lymphomas (TCLs) are less common.

Among the strongest risk factors for NHL are medical conditions, including those associated with immune dysfunction and chronic infections. For example, immunosuppression due to human immunodeficiency virus (HIV) infection or solid organ transplantation greatly increases NHL risk (4,5). Immunosuppression facilitates activation of Epstein-Barr virus infection, which contributes especially to DLBCL. Autoimmune diseases such as rheumatoid arthritis, systemic lupus erythematosus, and Sjögren syndrome increase risk for DLBCL and MZL (6). These medical conditions are thought to promote the development of NHL by causing long-term immune activation. Treatment of these conditions by immunosuppressive medications also likely contributes. In contrast, medical risk factors for CLL/SLL and FL are not clearly established.

Large administrative databases provide a valuable resource for examining associations between medical risk factors and cancer. We have previously used the SEER-Medicare database (described below) to conduct case-control studies among the US elderly population, assessing associations between various medical conditions and cancers such as NHL, leukemias, and skin cancers (7–10). Strengths of SEER-Medicare include availability of data from cancer registries (which provide reliable case ascertainment and detailed information on cancer subtypes), its large size (i.e., 1.6 million cancer cases diagnosed in 1991–2009), and information on medical risk factors detailed in Medicare claims beginning at age 65 (11). Until now, SEER-Medicare has been used to evaluate a limited number of medical risk factors for cancer (7–10,12–15), selected based on previously published findings or plausible biological mechanisms.

A comprehensive assessment of medical conditions in association with cancer would be an attractive new approach to characterizing a wide spectrum of risk factors. Using claims data, for example, this could be done by separately evaluating associations with every medical condition specified by billing codes. We term this novel approach "MedWAS," for "medical condition-wide association study," given its use of a broadly agnostic assessment that is a feature of genome-wide association studies (GWAS). One might anticipate that a MedWAS assessment of all medical conditions could uncover previously unsuspected associations with cancer, which would then prompt an investigation into possible biological mechanisms. A comprehensive assessment might also uncover a diversity of conditions associated with different cancer subtypes, providing evidence for etiologic heterogeneity.

In the present study, we implement this new MedWAS approach using SEER-Medicare data to assess potential risk factors for five subtypes of NHL. We demonstrate the utility of this method for characterizing the spectrum of medical risk factors for NHL, assessing etiologic heterogeneity among NHL subtypes, and identifying new medical conditions associated with NHL. Given the large multiplicity of testing, we emphasize the exploratory nature of this approach. We review possible etiologic and artifactual explanations for associations that we uncover.

Materials and Methods

Subject selection and ascertainment of medical conditions

SEER-Medicare (http://healthcaredelivery.cancer.gov/seermedicare/) links data from Surveillance, Epidemiology, and End Results (SEER) cancer registries (covering 28% of the US population in 2010) and Medicare (which provides medical insurance for the US elderly) (11). For this study, we selected cases and controls from the SEER-Medicare dataset as described previously (16). Specifically, we identified NHL cases in SEER that were indicated by SEER to be the person's first invasive cancer (except for a possible diagnosis of basal and squamous cell skin cancers, which are common non-melanoma skin cancers not captured by SEER). We included the five most common NHL subtypes, defined according to the World Health Organization classification (17): CLL/SLL, DLBCL, FL, MZL, and TCLs considered as a group.

All Medicare beneficiaries are enrolled in part A, which covers hospital care, and most also subscribe to part B which covers physician and outpatient services. Health maintenance organizations (HMOs) do not routinely bill Medicare for individual encounters. To ensure availability of Medicare claims prior to NHL diagnosis, we required that cases: 1) were age 66–99 years at diagnosis; 2) were diagnosed in 1992–2009; 3) had a minimum of 13 months of part A and part B Medicare coverage before diagnosis, during which they were not enrolled in an HMO; and 4) had at least one Medicare claim for a hospitalization (documented in the MEDPAR file), provider visit (NCH file), or outpatient services (OUTPATIENT file) at least 13 months before diagnosis. Medicare coverage and claims were considered back to the later of age 65 or a calendar year cutoff that varied according to calendar year of NHL diagnosis based on the availability of Medicare claims data (1991 for cases diagnosed in 1992–2002, 1998 for 2003–2005, 2000 for 2006–2007, 2002 for 2008–2009). Cases diagnosed only on autopsy or death certificate were excluded. Selection

yielded N=52,691 cases. For comparison, omission of the requirements for non-HMO coverage and at least 1 Medicare claim would have yielded N=68,044 cases.

Controls were selected from the random 5% sample of Medicare beneficiaries living in SEER areas included in the SEER-Medicare dataset (16). Controls were selected separately for each calendar year 1992–2009. As of July 1 (the selection date) of each year, controls were required to: 1) be alive and cancer-free; 2) have at least 13 months of prior Medicare coverage; and 3) have at least one Medicare claim at least 13 months earlier; age and calendar year cutoffs, and requirements for Medicare coverage and claims, were as described above for cases. From eligible controls, we randomly selected 200,000 controls, frequency matched to cases according to calendar year, sex, age, and race. Controls could be selected more than once for multiple years or included later as a case.

We searched Medicare claims to identify medical conditions diagnosed more than 12 months before case/control selection. The one-year period immediately before case/control selection was excluded to minimize bias due to reverse causality or differential medical work-up of cases. We initially considered medical conditions defined by the first four digits of International Classification of Diseases (version 9, ICD-9) codes. However, we also considered three-digit codes when providers only indicated this level of detail. To indicate that the condition was present, we required one inpatient claim with the diagnosis (MEDPAR file) or at least two physician or outpatient claims at least 30 days apart (NCH and OUTPATIENT files). Medical conditions could be described at any position in the claim, i.e., as primary or secondary diagnoses.

Statistical analysis

ICD-9 is a hierarchical coding system designed to provide an international standard for morbidity and mortality statistics, and especially as implemented in Medicare claims, for use in reimbursement of care providers. One challenge is that no level of the scheme uniformly captures all medical conditions at the same degree of detail, and some conditions are indicated by multiple codes in separate parts of the classification. We therefore used a step-wise approach to identify medical conditions associated with NHL subtypes (Figure 1).

Specifically, in the first step, the prevalence of every ICD-9-specified condition was compared separately between each NHL subtype and all controls. This group of unselected conditions (SELECT0) was defined by categorizing Medicare claims based on all provided four-digit ICD-9 codes (or occasionally, as noted above, by three-digit codes). In comparing the prevalence of the conditions in cases and controls, we selected conditions for further evaluation if: 1) the lowest achievable significance level computed from marginal totals in the 2×2 table (minalpha statistic) was less than 0.001 (18); and 2) the p-value from the Cochran–Mantel–Haenszel test (conditioning on the matching factors) was less than the Bonferroni cutoff (defined as 0.05 divided by the number of conditions remaining after applying the minalpha criterion). We excluded ICD-9-coded conditions for invalid conditions, specifically those obviously corresponding to a possible NHL diagnosis (e.g., NHL itself, lymphadenopathy, or splenomegaly), non-specific symptoms (e.g., headache, fatigue), and spurious codes that could not be matched to diagnoses. Because cases and controls were selected to have no prior SEER-documented cancer, we considered claims for

previous cancer diagnoses (other than non-melanoma skin cancer) as having uncertain reliability; therefore, we also excluded these claims diagnoses from analysis.

This procedure yielded a subset of conditions for each NHL subtype, which we refer to as SELECT1. Next, we used binary logistic regression models to derive odds ratios (ORs), measuring the associations of each SELECT1 condition with the NHL subtype, adjusted for demographic characteristics (sex, age and calendar year of case/control selection, race), and as a measure of healthcare utilization, the number of provider claims per year (see Table 1 footnote for details). Each NHL subtype was compared with all controls, and the variance of the ORs accounted for the multiple sampling of some controls (16).

Two physicians reviewed these results to group similar SELECT1 conditions together (e.g., multiple ICD-9 codes for skin cancer at different body sites), add related ICD-9 codes not in SELECT1 as part of the groups, and in rare instances, to break SELECT1 conditions into finer categories (e.g., distinguishing different hepatitis virus infections based on five-digit codes). SELECT1 conditions were removed if they were rare (<5 affected cases when the OR>1, or <100 affected controls when the OR<1) and could not be grouped with other conditions. This process led to SELECT2 conditions for each NHL subtype.

We again used binary logistic regression to assess associations of SELECT2 conditions, adjusted for demographic characteristics and yearly physician claims. We excluded SELECT2 conditions for which the adjusted OR was both non-significant (p 0.05) and close to the null value, or for which the models did not converge due to small numbers of affected cases or controls. This process yielded the final group of SELECT3 medical conditions associated with each NHL subtype.

We compiled the list of SELECT3 medical conditions across the five NHL subtypes and used polytomous logistic regression to assess the association of each SELECT3 condition with each subtype, adjusting for demographic characteristics and yearly provider claims. Although we present ORs for each condition with all five subtypes, we focus on the subset of conditions that were associated with each subtype in its separate SELECT3 analysis. Each polytomous logistic regression model provided a test of heterogeneity of the ORs across NHL subtypes.

We also conducted a sensitivity analysis for SELECT3 conditions to further minimize the possibility of reverse causality. Specifically, we excluded from evaluation the three-year period immediately preceding NHL diagnosis/control selection. This approach left N=39,995 cases and N=158,706 controls (age 68 or older) with evaluable time covered by Medicare. For these subjects, we re-ascertained the previously identified SELECT3 medical conditions, this time excluding the three-year window before case/control selection, and reran the polytomous logistic regression models.

Replication of selected findings

We sought to replicate selected associations with SELECT3 medical conditions in two independent datasets: the National Institutes of Health-AARP Diet and Health Study (NIH-AARP) and the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO), as

described in the notes to Supplemental Table 1. These replications required data on both the medical condition and NHL subtypes, so replications could not be undertaken for all findings. We utilized a two-sided alpha of 0.05 to assess significance in these analyses. In some instances, the magnitude of the association that we tried to replicate and the number of cases with each NHL subtype were both small, greatly limiting the statistical power for the replication (Supplemental Table 1).

Results

Characteristics of study subjects and stepwise selection of medical conditions

We included 52,691 cases with NHL (N=19,078 DLBCL, N=18,236 CLL/SLL, N=8881 FL, N=4289 MZL, N=2207 TCL) from SEER-Medicare. Overall, cases were well-matched to the 200,000 controls according to demographic characteristics, although there were some minor differences among NHL subtypes (Table 1). Cases tended to have slightly shorter duration of prior Medicare coverage (median 52 vs. 54 months, excluding the 12 months immediately before case/control selection) but had more physician visits per year (Table 1).

Figure 1 and Table 2 document the process by which medical conditions associated with each NHL subtype were identified. For each NHL subtype, we screened 5605 to 5785 conditions indicated by unique four-digit (and occasional three-digit) ICD-9 codes, or a total of 5926 conditions across all five subtypes (SELECT0 conditions). More than 97% of these conditions were excluded because they were not significantly associated with the subtype or were for invalid diagnoses. This procedure left 30 to 52 remaining conditions for each subtype, identified by (mostly) four-digit ICD-9 codes SELECT1 conditions (see Supplemental Table 2 for a complete list). Two physicians reviewed these conditions and grouped related codes to create 13 to 28 SELECT2 conditions for each NHL subtype.

Most SELECT2 conditions remained associated with their respective NHL subtypes in multivariate logistic regression models, yielding the final SELECT3 group of medical conditions (N=18 conditions for CLL/SLL, N=27 for DBLCL, N=13 for FL, N=18 for MZL, and N=15 for TCL, for a total of N=55 unique SELECT3 conditions; Table 2). ICD-9 codes for SELECT3 conditions, as well as the number of subjects with each condition, are presented in Supplemental Table 3.

Findings of MedWAS analyses

Table 3 presents associations for the 55 SELECT3 medical conditions with each NHL subtype. Of note, most conditions (N=49, 89%) varied significantly in their associations across NHL subtypes (i.e., p-heterogeneity<0.05; Table 3).

We highlight associations for each subtype that were identified in the SELECT3 group for that subtype (shaded in gray in Table 3). For most of these, the 95% confidence interval for the odds ratio (OR) excludes 1.00 (underlined in Table 3). Only 3 of the 55 medical conditions were associated with increased risk for all five NHL subtypes: non-melanoma skin cancer (ORs 1.19–1.55), actinic keratosis (1.12–1.25), and hemolytic anemia (1.64–4.07) (Table 3).

Among the positive associations, several immunodeficiency or infectious conditions were associated with increased risk for 1–2 NHL subtypes (Table 3). Specifically, associations were observed for HIV infection and solid organ transplantation with DLBCL (ORs 3.83 and 4.27, respectively), and deficiency of humoral immunity with CLL/SLL (3.85) and TCL (5.70). CLL/SLL risk was increased in association with herpes zoster (OR 1.44), acute sinusitis (1.12), and acute bronchitis (1.06), and DLBCL was increased with hepatitis C virus (HCV) infection (1.74).

Autoimmune diseases were also associated with increased risk for some subtypes, including for DLBCL with rheumatoid arthritis (OR 1.43), sarcoidosis (2.11), and uveitis (3.17). Systemic lupus erythematosus was associated with risk of DLBCL (OR 1.74) and MZL (2.57); Sjögren syndrome with DLBCL (2.10) and MZL (4.74); and celiac disease with TCL (8.09). An elevated erythrocyte sedimentation rate was associated with increased risk of MZL (OR 2.00).

Among hematologic conditions (in addition to hemolytic anemia which was associated with all NHL subtypes), thrombocytopenia was associated with increased risk for all subtypes other than FL (ORs 1.46–2.21). Aplastic anemia, anemia not otherwise specified (NOS), and neutropenia were positively associated with MZL (ORs 3.26, 1.12, and 2.47, respectively). Monoclonal paraproteinemia was associated with CLL/SLL, DLBCL, and MZL (ORs 1.80–3.43), and cryoglobulinemia with DLBCL (6.36) and FL (10.14).

Among skin conditions other than non-melanoma skin cancer and actinic keratosis, nine were associated with increased risk only for TCL: atopic dermatitis (OR 4.12), contact dermatitis (2.61), dermatitis due to substances taken internally (4.05), bullous skin diseases (3.43), discoid lupus (4.00), psoriasis (3.72), folliculitis (2.65), asteatosis (2.20), and urticaria (2.42). Seborrheic keratosis was associated with increased risk of FL (OR 1.18) and MZL (1.24).

There were a few additional positive associations. Diabetes mellitus was associated with DLBCL (OR 1.09). Testicular hypofunction, benign prostatic hyperplasia, and gastric ulcer were associated with MZL (ORs 1.58, 1.30, and 1.55, respectively), and spinal cord anomalies was associated with CLL/SLL (3.20).

No medical condition was inversely associated with all five NHL subtypes. However, inverse associations were observed with several neurological/psychiatric conditions (Table 3). Three such conditions were inversely associated with all NHL subtypes except TCL: senile dementia (ORs 0.40–0.61), stroke (ORs 0.74–0.86), and psychosis NOS (0.37–0.64). Inverse associations were also observed for arteriosclerotic dementia with CLL/SLL, DLBCL, and FL (ORs 0.32–0.54); Parkinson's disease with DLBCL (0.67); depression NOS with DLBCL (0.74) and FL (0.70); non-psychotic mental disorder NOS with DLBCL (0.23); and alcoholism with DLBCL (0.59).

Among infections, inverse associations were found for chronic bronchitis and urinary tract infection with DLBCL (ORs 0.78 and 0.83, respectively). CLL/SLL was reduced in association with hypertension (OR 0.83), abdominal aortic aneurysm (0.67), and hyperlipidemia (0.86), and FL was reduced in association with systolic heart failure (0.80).

Among miscellaneous conditions, inverse associations were observed between decubitus ulcer and DLBCL (OR 0.57); asphyxia and FL (0.52); and hip fracture and DLBCL (0.62) and FL (0.50).

Sensitivity and replication analyses

We performed a sensitivity analyses for the 55 SELECT3 medical conditions, excluding claims during the three years immediately before case/control selection. ORs were very similar to the primary analysis (Supplemental Table 4).

Table 4 presents results of replication analyses in additional populations for some SELECT3 conditions. For diabetes mellitus and DLBCL, positive associations in the NIH-AARP and PLCO cohorts appeared consistent with the MedWAS observation, although the replications did not reach statistical significance. There was also an inverse (though statistically non-significant) association in PLCO between chronic bronchitis and DLBCL. Associations with hypertension and stroke were not significant in replication analyses.

Discussion

We surveyed a large number of medical conditions as risk factors for NHL using a new approach termed "MedWAS." This method characterized the full spectrum of medical conditions related to NHL among the US elderly. Some associations that we document are well established, while others are new or less supported..

Importantly, for the 55 medical conditions retained in our final (SELECT3) analyses, we found that most associations varied significantly across the five NHL subtypes. In fact, most medical conditions were associated with only one or a few NHL subtypes. Because some associations are likely etiologic (as we review below), these differences point to etiologic heterogeneity, i.e., distinct causal factors that contribute to each NHL subtype. Other differences in associations with environmental risk factors and genetic polymorphisms (3,19–22) likewise support that different NHL subtypes arise through separate (although perhaps overlapping) mechanisms.

Some associations that we demonstrated reflect well-established contributions of chronic immune disturbances to development of NHL. These include associations of HIV infection, HCV infection, and solid organ transplantation with DLBCL and, to a lesser extent, MZL (4,5,23). As described previously (6), we found an increased risk of DLBCL and/or MZL associated with a range of autoimmune conditions including rheumatoid arthritis, systemic lupus erythematosus, and Sjögren syndrome, and strongly elevated risk of TCL with celiac disease. In striking contrast, none of these conditions was associated with CLL/SLL or FL. Increased risk of CLL/SLL following herpes zoster, acute sinusitis, and acute bronchitis may plausibly be a manifestation of chronic immune deficits preceding this malignancy (24,25).

Interestingly, risk for each NHL subtype was elevated following non-melanoma skin cancer. A few prior studies have described increased risk for NHL or CLL/SLL following diagnoses of basal or squamous cell skin cancers (26–28), but the present investigation is the first to document associations for specific NHL subtypes other than CLL/SLL. This study is also

the first to show an increased risk for NHL following a diagnosis of actinic keratosis, the precursor of squamous cell skin cancer. Although skin damage from ultraviolet radiation strongly increases risk for skin cancer, ultraviolet radiation actually appears inversely associated with NHL (29). More likely, the association of non-melanoma skin cancer and NHL is related to immunosuppression, as suggested by associations of HIV and solid organ transplantation with skin cancer (30,31).

Hematologic conditions associated with increased NHL risk in our study have been described as complications of lymphoproliferative conditions (e.g., hemolytic anemia and cryoglobulinemia) or non-specific manifestations of chronic illness (32–35). It is unlikely that these conditions were caused by undiagnosed NHL (i.e., reverse causality), because we did not consider Medicare claims within one year before NHL diagnosis, and remarkably, almost all of the associations persisted in sensitivity analyses excluding claims within three years of NHL. Instead, these associations may again reflect the presence of chronic immune disturbances that contribute to the development of NHL over a prolonged period.

We found strong associations between a large number of dermatologic conditions and risk of TCL, some of which have been observed previously (6,36). TCLs can have an indolent presentation (37), and 48% of TCLs in our study were cutaneous lymphomas (mycosis fungoides and less common variants). Diagnostic confusion between skin conditions and TCL could be an explanation, but the associations persisted when we excluded diagnoses within three years of NHL. Alternatively, these associations may reflect a shared predisposition to skin diseases and TCL, immune effects of chronic skin diseases, or effects of treatment of the skin diseases (38) .

The association of gastric ulcer with MZL may be explained by diagnostic confusion between gastric ulcers and gastric MZL, or by the etiologic contribution of *Helicobacter pylori* to both conditions (39,40). There was also a positive association of diabetes mellitus with DLBCL. This finding showed some evidence, although inconclusive, for replication in our analyses in NIH-AARP and PLCO. Diabetes mellitus has previously been associated with increased risk of NHL overall (41,42).

Some inverse associations with NHL risk could have biological explanations. Decreased risk of CLL/SLL following a diagnosis of hyperlipidemia is intriguing, and prior studies have noted a protective effect of statins (a widely used class of lipid-lowering medications) for NHL overall and leukemia (43). Decreased CLL/SLL risk associated with alcoholism may reflect protective effects of ethanol consumption (3).

Given our lack of data on lifestyle factors, it is possible that some associations could be due to confounding, e.g., by smoking, drinking, or occupation (3). Other artifacts could underlie inverse associations. It is likely that clinicians limited the medical work-up of some frail and debilitated elderly adults among the Medicare population, which would have led to under-ascertainment of NHL. This bias could explain the inverse associations with a broad range of neurological and psychiatric conditions and conditions associated with advanced illness or nursing home care (e.g., decubitus ulcer, hip fracture). Indeed, we did not observe decreased NHL risk when, in our replication analyses, we assessed people with a history of a stroke in

the NIH-AARP and PLCO studies. Participants in these cohort studies were younger and would have been healthier than unselected Medicare beneficiaries, and so NHLs arising in these individuals would have been less vulnerable to under-diagnosis.

The present study is the first implementation of a new method, "MedWAS," which we used to comprehensively evaluate a very large number of medical conditions as NHL risk factors. Demonstration of multiple known associations with NHL supports the validity of this approach. MedWAS incorporates the same agnostic, wide-based approach used in GWAS studies to survey thousands of DNA variations (44). MedWAS could also be applied to other sources of administrative data and electronic health records. Recent "big data" analyses of large administrative databases by others have focused on characterizing the network properties of related medical conditions (45) or, with respect to cancer, selecting optimum patient treatments and predicting outcomes (46).

Several limitations of our study should be noted. First, it was restricted to the US elderly, and our results may not generalize to other populations. Although Medicare covers essentially all US adults over age 65, our requirements that subjects were not in an HMO and had at least one documented claim led to some exclusions. Further, we assessed Medicare data beginning at age 65, so we were unable to evaluate medical conditions that did not generate claims at older ages. Second, we are unaware of a systematic method for using ICD-9 codes to classify unique medical conditions at a consistent and informative level of detail. We therefore found it necessary to review individual codes to identify biologically relevant conditions and to eliminate invalid or irrelevant codes, which likely introduced some subjectivity. Medicare claims can be inaccurate, but we sought to increase the positive predictive value by requiring one inpatient or two physician/outpatient claims at least 30 days apart (16). Also, it is not possible to assess duration or severity of medical conditions using Medicare claims, and as our study was exploratory, we did not attempt to examine associations with treatments for these conditions. Third, because we made thousands of comparisons, some associations could have been due to chance. We sought to minimize this issue by requiring strict statistical significance in the initial screening of the ICD-9 codes and by attempting to independently replicate some findings. Unfortunately, it was challenging to find appropriate data sources for replication, and due to the small number of outcomes and modest size of the associations, the replication analyses were inconclusive.

In conclusion, our study comprehensively assessed a very large number of medical conditions and thereby identified a subset associated with increased or decreased NHL risk. Specific associations varied according to NHL subtype. Many risk factors were related to immune disturbances and chronic infections, as expected, but some (such as the associations with non-melanoma skin cancer, skin conditions, and diabetes) point to new avenues for research. It will be important to replicate some findings in additional populations, and to uncover biological mechanisms underpinning the best supported and strongest associations. We believe that this MedWAS approach can be useful in epidemiologic research aimed at understanding the etiology of cancer and other complex diseases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Reference List

1. Fitzmaurice C, Dicker D, Pain A, Hamavid H, Moradi-Lakeh M, MacIntyre MF, et al. The Global Burden of Cancer 2013. JAMA Oncol. 2015; 1:505–527. [PubMed: 26181261]

2. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 18 Registries Research Data. 2013 Nov. Submission (2000–2011)

3. Morton LM, Slager SL, Cerhan JR, Wang SS, Vajdic CM, Skibola CF, et al. Etiologic heterogeneity among non-Hodgkin lymphoma subtypes: the InterLymph Non-Hodgkin Lymphoma Subtypes Project. J Natl Cancer Inst Monogr. 2014; 2014:130–144. [PubMed: 25174034]

4. Gibson TM, Morton LM, Shiels MS, Clarke CA, Engels EA. Risk of non-Hodgkin lymphoma subtypes in HIV-infected people during the HAART era: a population-based study. AIDS. 2014; 28:2313–2318. [PubMed: 25111081]

5. Clarke CA, Morton LM, Lynch C, Pfeiffer RM, Hall EC, Gibson TM, et al. Risk of lymphoma subtypes after solid organ transplantation in the United States. Br J Cancer. 2013; 109:280–288. [PubMed: 23756857]

6. Ekstrom Smedby K, Vajdic CM, Falster M, Engels EA, Martinez-Maza O, Turner J, et al. Autoimmune disorders and risk of non-Hodgkin lymphoma subtypes: a pooled analysis within the InterLymph Consortium. Blood. 2008; 111:4029–4038. [PubMed: 18263783]

7. Anderson LA, Gadalla S, Morton LM, Landgren O, Pfeiffer R, Warren JL, Berndt SI, Ricker W, Parsons R, Engels EA. Population-based study of autoimmune conditions and the risk of specific lymphoid malignancies. Int J Cancer. 2009; 125:398–405. [PubMed: 19365835]

8. Anderson LA, Pfeiffer RM, Landgren O, Gadalla S, Berndt SI, Engels EA. Risks of myeloid malignancies in patients with autoimmune conditions. Br J Cancer. 2009; 100:822–828. [PubMed: 19259097]

9. Lanoy E, Costagliola D, Engels EA. Skin cancers associated with HIV infection and solid-organ transplantation among elderly adults. Int J Cancer. 2010; 126:1724–1731. [PubMed: 19810102]

10. Lanoy E, Engels EA. Skin cancers associated with autoimmune conditions among elderly adults. Br J Cancer. 2010; 103:112–114. [PubMed: 20551958]

11. Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. Med Care. 2002; 40 IV-18.

12. Welzel TM, Graubard BI, El-Serag HB, Shaib YH, Hsing AW, Davila JA, et al. Risk factors for intrahepatic and extrahepatic cholangiocarcinoma in the United States: a population-based case-control study. Clin Gastroenterol Hepatol. 2007; 5:1221–1228. [PubMed: 17689296]

13. Welzel TM, Graubard BI, Zeuzem S, El-Serag HB, Davila JA, McGlynn KA. Metabolic syndrome increases the risk of primary liver cancer in the United States: a study in the SEER-Medicare database. Hepatology. 2011; 54:463–471. [PubMed: 21538440]

14. Trabert B, Wentzensen N, Felix AS, Yang HP, Sherman ME, Brinton LA. Metabolic syndrome and risk of endometrial cancer in the united states: a study in the SEER-medicare linked database. Cancer Epidemiol Biomarkers Prev. 2015; 24:261–267. [PubMed: 25587111]

15. Nogueira L, Freedman ND, Engels EA, Warren JL, Castro F, Koshiol J. Gallstones, cholecystectomy, and risk of digestive system cancers. Am J Epidemiol. 2014; 179:731–739. [PubMed: 24470530]

16. Engels EA, Pfeiffer RM, Ricker W, Wheeler W, Parsons R, Warren JL. Use of surveillance, epidemiology, and end results-medicare data to conduct case-control studies of cancer among the US elderly. Am J Epidemiol. 2011; 174:860–870. [PubMed: 21821540]

17. Turner JJ, Morton LM, Linet MS, Clarke CA, Kadin ME, Vajdic CM, et al. InterLymph hierarchical classification of lymphoid neoplasms for epidemiologic research based on the WHO classification (2008): update and future directions. Blood. 2010; 116:e90–e98. [PubMed: 20699439]

18. Tarone RE. A modified Bonferroni method for discrete data. Biometrics. 1990; 46:515–522. [PubMed: 2364136]

19. Morton LM, Wang SS, Cozen W, Linet MS, Chatterjee N, Davis S, et al. Etiologic heterogeneity among non-Hodgkin lymphoma subtypes. Blood. 2008; 112:5150–5160. [PubMed: 18796628]

20. Berndt SI, Skibola CF, Joseph V, Camp NJ, Nieters A, Wang Z, et al. Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. Nat Genet. 2013; 45:868–876. [PubMed: 23770605]

21. Cerhan JR, Berndt SI, Vijai J, Ghesquieres H, McKay J, Wang SS, et al. Genome-wide association study identifies multiple susceptibility loci for diffuse large B cell lymphoma. Nat Genet. 2014; 46:1233–1238. [PubMed: 25261932]

22. Skibola CF, Berndt SI, Vijai J, Conde L, Wang Z, Yeager M, et al. Genome-wide association study identifies five susceptibility loci for follicular lymphoma outside the HLA region. Am J Hum Genet. 2014; 95:462–471. [PubMed: 25279986]

23. de Sanjose S, Benavente Y, Vajdic CM, Engels EA, Morton LM, Bracci PM, et al. Hepatitis C and non-Hodgkin lymphoma among 4784 cases and 6269 controls from the International Lymphoma Epidemiology Consortium. Clin Gastroenterol Hepatol. 2008; 6:451–458. [PubMed: 18387498]

24. Anderson LA, Landgren O, Engels EA. Common community acquired infections and subsequent risk of chronic lymphocytic leukaemia. Br J Haematol. 2009

25. Landgren O, Gridley G, Check D, Caporaso NE, Morris BL. Acquired immune-related and inflammatory conditions and subsequent chronic lymphocytic leukaemia. Br J Haematol. 2007; 139:791–798. [PubMed: 17941950]

26. Levi F, Randimbison L, Te VC, La Vecchia C. Non-Hodgkin's lymphomas, chronic lymphocytic leukaemias and skin cancers. Br J Cancer. 1996; 74:1847–1850. [PubMed: 8956805]

27. Wheless L, Black J, Alberg AJ. Nonmelanoma skin cancer and the risk of second primary cancers: a systematic review. Cancer Epidemiol Biomarkers Prev. 2010; 19:1686–1695. [PubMed: 20570907]

28. Adami J, Frisch M, Yuen J, Glimelius B, Melbye M. Evidence of an association between non-Hodgkin's lymphoma and skin cancer. BMJ. 1995; 310:1491–1495. P. [PubMed: 7787593]

29. Kricker A, Armstrong BK, Hughes AM, Goumas C, Smedby KE, Zheng T, et al. Personal sun exposure and risk of non Hodgkin lymphoma: a pooled analysis from the Interlymph Consortium. Int J Cancer. 2008; 122:144–154. [PubMed: 17708556]

30. Euvrard S, Kanitakis J, Claudy A. Skin cancers after organ transplantation. N Engl J Med. 2003; 348:1681–1691. [PubMed: 12711744]

31. Silverberg MJ, Leyden W, Warton EM, Quesenberry CP Jr, Engels EA, Asgari MM. HIV infection status, immunodeficiency, and the incidence of non-melanoma skin cancer. J Natl Cancer Inst. 2013; 105:350–360. [PubMed: 23291375]

32. Ding W, Zent CS. Diagnosis and management of autoimmune complications of chronic lymphocytic leukemia/ small lymphocytic lymphoma. Clin Adv Hematol Oncol. 2007; 5:257–261. [PubMed: 17607284]

33. Zent CS, Kay NE. Autoimmune complications in chronic lymphocytic leukaemia (CLL). Best Pract Res Clin Haematol. 2010; 23:47–59. [PubMed: 20620970]

34. Molina-Garrido MJ, Guillen-Ponce C. A revision on cryoglobulinaemia associated to neoplastic diseases. Clin Transl Oncol. 2007; 9:229–236. [PubMed: 17462975]

35. Murakami H, Irisawa H, Saitoh T, Matsushima T, Tamura J, Sawamura M, et al. Immunological abnormalities in splenic marginal zone cell lymphoma. Am J Hematol. 1997; 56:173–178. [PubMed: 9371530]

36. Legendre L, Barnetche T, Mazereeuw-Hautier J, Meyer N, Murrell D, Paul C. Risk of lymphoma in patients with atopic dermatitis and the role of topical treatment: A systematic review and meta-analysis. J Am Acad Dermatol. 2015; 72:992–1002. [PubMed: 25840730]

37. Willemze R, Jaffe ES, Burg G, Cerroni L, Berti E, Swerdlow SH, et al. WHO-EORTC classification for cutaneous lymphomas. Blood. 2005; 105:3768–3785. [PubMed: 15692063]

38. Aschebrook-Kilfoy B, Cocco P, La Vecchia C, Chang ET, Vajdic CM, Kadin ME, et al. Medical history, lifestyle, family history, and occupational risk factors for mycosis fungoides and Sezary syndrome: the InterLymph Non-Hodgkin Lymphoma Subtypes Project. J Natl Cancer Inst Monogr. 2014; 2014:98–105. [PubMed: 25174030]

39. Malfertheiner P, Link A, Selgrad M. Helicobacter pylori: perspectives and time trends. Nat Rev Gastroenterol Hepatol. 2014; 11:628–638. [PubMed: 25001975]

40. Dover F, Ipek S. Malignancy risk of gastric ulcers: could it be higher than the expected values? Hepatogastroenterology. 2003; (50 Suppl 2):cccxii–cccxiv. [PubMed: 15244212]

41. Mitri J, Castillo J, Pittas AG. Diabetes and risk of Non-Hodgkin's lymphoma: a meta-analysis of observational studies. Diabetes Care. 2008; 31:2391–2397. [PubMed: 19033419]

42. Chao C, Page JH. Type 2 diabetes mellitus and risk of non-Hodgkin lymphoma: a systematic review and meta-analysis. Am J Epidemiol. 2008; 168:471–480. [PubMed: 18611954]

43. Pradelli D, Soranna D, Zambon A, Catapano A, Mancia G, La Vecchia C, et al. Statins use and the risk of all and subtype hematological malignancies: a meta-analysis of observational studies. Cancer Med. 2015; 4:770–780. PMCID: PMC4430269. [PubMed: 25809667]

44. Chung CC, Chanock SJ. Current status of genome-wide association studies in cancer. Hum Genet. 2011; 130:59–78. [PubMed: 21678065]

45. Hidalgo CA, Blumm N, Barabasi AL, Christakis NA. A dynamic network approach for the study of human phenotypes. PLoS Comput Biol. 2009; 5:e1000353. [PubMed: 19360091]

46. Kantarjian H, Yu PP. Artificial Intelligence, Big Data, and Cancer. JAMA Oncol. 2015; 1:573–574. [PubMed: 26181906]
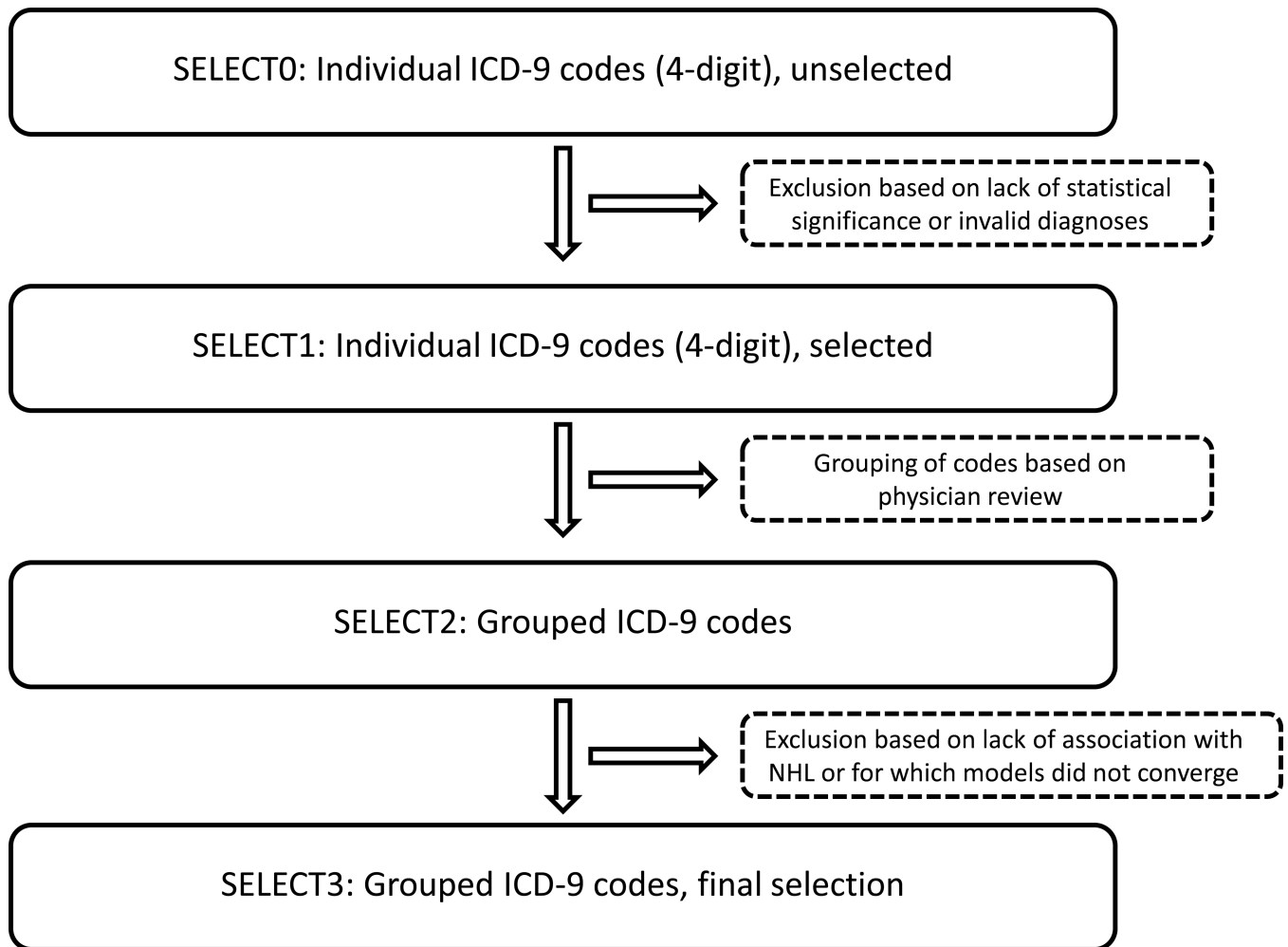
Figure 1.
Stepwise selection of medical conditions associated with subtypes of non-Hodgkin
lymphoma. The figure illustrates the steps used to identify a final group of medical
conditions, defined by ICD-9 codes, that were associated with each NHL subtype.

Table 1

Characteristics of non-Hodgkin lymphoma cases and controls in SEER-Medicare (1992–2009)

Characteristic, N (%)	Controls (N=200,000)	NHL cases combined (N=52,691)	CLL/SLL (N=18,236)	DLBCL (N=19,078)	FL (N=8881)	MZL (N=4289)	TCL (N=2207)
Sex							
Male	95,262 (47.6)	25,096 (47.6)	9675 (53.1)	8640 (45.3)	3828 (43.1)	1748 (40.8)	1205 (54.6)
Female	104,738 (52.4)	27,595 (52.4)	8561 (47.0)	10438 (54.7)	5053 (56.9)	2541 (59.2)	1002 (45.4)
Age in years at case/control selection [*]							
65–69	30,397 (15.2)	8008 (15.2)	2688 (14.7)	2578 (13.5)	1685 (19.0)	691 (16.1)	366 (16.6)
70–74	48,228 (24.1)	12,706 (24.1)	4242 (23.3)	4425 (23.2)	2404 (27.1)	1025 (23.9)	610 (27.6)
75–79	49,090 (24.6)	12,934 (24.6)	4430 (24.3)	4784 (25.1)	2168 (24.4)	1012 (23.6)	540 (24.5)
80–84	39,579 (19.8)	10,428 (19.8)	3531 (19.4)	3973 (20.8)	1622 (18.3)	902 (21.0)	400 (18.1)
85+	32,706 (16.4)	8615 (16.4)	3345 (18.3)	3318 (17.4)	1002 (11.3)	659 (15.4)	291 (13.2)
Race							
White	178,689 (89.3)	47,080 (89.4)	16,679 (91.5)	16,701 (87.5)	8160 (91.9)	3751 (87.5)	1789 (81.1)
Black	9267 (4.6)	2439 (4.6)	1027 (5.6)	694 (3.6)	292 (3.3)	208 (4.9)	218 (9.9)
Other	12,044 (6.0)	3172 (6.0)	530 (2.9)	1683 (8.8)	429 (4.8)	330 (7.7)	200 (9.1)
Duration of Medicare coverage in months at case/control selection [*]							
1–24	37,880 (18.9)	11,936 (22.7)	4352 (23.9)	4101 (21.5)	2184 (24.6)	803 (18.7)	496 (22.5)
25–49	50,070 (25.0)	12,759 (24.2)	4612 (25.3)	4441 (23.3)	2194 (24.7)	966 (22.5)	546 (24.7)
50–74	62,543 (31.3)	17,681 (33.6)	5898 (32.3)	6533 (34.2)	2885 (32.5)	1620 (37.8)	745 (33.8)
75+	49,507 (24.8)	10,315 (19.6)	3374 (18.5)	4003 (21.0)	1618 (18.2)	900 (21.0)	420 (19.0)
Average number of physician visits per year prior to case/control selection [†]							
0	56,079 (28.0)	2921 (5.5)	1143 (6.3)	1065 (5.6)	483 (5.4)	107 (2.5)	123 (5.6)
0.001–3.124	43,807 (21.9)	12,862 (24.4)	4413 (24.2)	4799 (25.2)	2265 (25.5)	926 (21.6)	459 (20.8)
3.125–7.499	49,969 (25.0)	17,304 (32.8)	5921 (32.5)	6248 (32.8)	2996 (33.7)	1413 (32.9)	726 (32.9)
7.500+	50,145 (25.1)	19,604 (37.2)	6759 (37.1)	6966 (36.5)	3137 (35.3)	1843 (43.0)	899 (40.7)

Author Manuscript

Abbreviations: NHL non-Hodgkin lymphoma, CLL/SLL chronic lymphocytic leukemia/small lymphocytic lymphoma, DLBCL diffuse large B-cell lymphoma, MZL marginal zone lymphoma, TCL T-cell lymphoma.

*
Medicare coverage refers to part A and part B coverage, without enrollment in a health maintenance organization. The coverage in the 12 months prior to NHL diagnosis/control selection is not counted in this total.

†
Physician visits are for the period excluding the 12 months prior to case/control selection. This variable is based on a count of records in the NCH file that have the following Healthcare Common Procedure Coding System (HCPCS) codes: 99201–99255, 99261–99263, 99271–99275,99301–99340, 99341–99353, 99354–99357, 99381–99429. The count excludes records for radiology, anesthesiology, and pathology providers, and includes a maximum of one NCH record per day.

Table 2

Number of evaluated medical conditions at each stage of analysis of SEER-Medicare claims data

Analysis Stage	Description	Number of conditions, by NHL subtype						Number of Unique Conditions across all subtypes
		CLL/SLL	DLBCL	FL	MZL	TCL		
SELECT0	Individual ICD-9 codes, unselected	5762	5785	5712	5618	5605		5926
SELECT1	Individual ICD-9 codes, selected for statistical significance according to multiple testing procedure, excluding invalid conditions	30	43	27	50	52		146
SELECT2	Grouped ICD-9 codes, based on review by physicians	17	28	13	18	18		59
SELECT3	Grouped ICD-9 codes, excluding conditions that were not associated with NHL subtype or for which models did not converge	18	27	13	18	15		55

Abbreviations: NHL non-Hodgkin lymphoma, CLL/SLL chronic lymphocytic leukemia/small lymphocytic lymphoma, DLBCL diffuse large B-cell lymphoma, MZL marginal zone lymphoma, TCL T-cell lymphoma, ICD-9 International Classification of Diseases version 9.

Table 3

Odds ratios associating SELECT3 medical conditions with NHL, SEER-Medicare (1992–2009)

Condition	Adjusted odds ratio (95%CI)*					Heterogeneity p-value
	CLL/SLL	DLBCL	FL	MZL	TCL	
IMMUNODEFICIENCY RELATED CONDITIONS						
HIV	0.86 (0.34–2.16)	3.83 (2.28–6.43)	0.40 (0.06–2.90)	2.15 (0.66–6.96)	0	<0.0001
Deficiency of humoral immunity	3.85 (2.43–6.11)	1.42 (0.72–2.81)	0.57 (0.14–2.31)	2.67 (1.05–6.75)	5.70 (2.29–14.18)	0.0058
Solid organ transplantation	0.91 (0.53–1.55)	4.27 (3.23–5.64)	0.95 (0.47–1.95)	1.92 (0.97–3.79)	3.58 (1.82–7.03)	<0.0001
INFECTIONS						
Herpes zoster	1.44 (1.28–1.61)	1.23 (1.10–1.39)	1.04 (0.86–1.25)	1.42 (1.15–1.76)	1.20 (0.85–1.68)	0.0244
HCV	0.61 (0.41–0.92)	1.74 (1.37–2.22)	0.62 (0.36–1.08)	1.27 (0.77–2.10)	0.62 (0.23–1.67)	<0.0001
Acute sinusitis	1.12 (1.05–1.20)	0.98 (0.92–1.05)	1.12 (1.02–1.22)	1.10 (0.97–1.24)	0.95 (0.79–1.15)	0.0170
Acute bronchitis	1.06 (1.01–1.11)	0.94 (0.90–0.99)	0.94 (0.88–1.02)	1.02 (0.92–1.12)	0.96 (0.84–1.11)	0.0069
Chronic bronchitis	0.82 (0.77–0.89)	0.78 (0.73–0.84)	0.79 (0.71–0.88)	0.95 (0.83–1.08)	0.73 (0.60–0.89)	0.0719
Urinary tract infection	0.89 (0.85–0.93)	0.83 (0.80–0.87)	0.83 (0.78–0.88)	0.79 (0.72–0.86)	0.77 (0.68–0.88)	0.0279
AUTOIMMUNE AND INFLAMMATORY CONDITIONS						
Rheumatoid arthritis	0.90 (0.82–1.00)	1.43 (1.32–1.55)	1.20 (1.07–1.36)	1.18 (1.01–1.39)	1.31 (1.05–1.64)	<0.0001
Systemic lupus erythematosus	0.82 (0.60–1.14)	1.74 (1.39–2.17)	1.06 (0.72–1.55)	2.57 (1.84–3.58)	2.25 (1.34–3.77)	<0.0001
Sjogren syndrome	1.05 (0.82–1.33)	2.10 (1.77–2.49)	1.57 (1.20–2.06)	4.74 (3.81–5.89)	1.78 (1.08–2.93)	<0.0001
Sarcoidosis	0.84 (0.50–1.39)	2.11 (1.50–2.96)	0.92 (0.47–1.81)	2.41 (1.37–4.24)	0.76 (0.19–3.07)	0.0033
Celiac disease	1.41 (0.84–2.34)	1.34 (0.81–2.23)	0.98 (0.43–2.23)	2.43 (1.19–4.96)	8.09 (4.36–15.02)	<0.0001
Uveitis	1.62 (0.85–3.09)	3.17 (1.97–5.08)	1.46 (0.58–3.64)	1.15 (0.28–4.73)	0	<0.0001
Elevated sedimentation rate	1.03 (0.79–1.36)	1.31 (1.03–1.66)	0.76 (0.49–1.19)	2.00 (1.39–2.86)	1.33 (0.68–2.58)	0.0074
HEMATOLOGIC CONDITIONS						
Hemolytic anemia	3.60 (2.70–4.82)	1.99 (1.39–2.85)	1.64 (0.94–2.85)	4.07 (2.51–6.60)	2.97 (1.37–6.41)	0.0041
Aplastic anemia	1.49 (1.21–1.82)	1.37 (1.12–1.68)	1.13 (0.82–1.58)	3.26 (2.51–4.22)	1.20 (0.68–2.13)	<0.0001
Anemia NOS	1.01 (0.97–1.06)	1.03 (0.98–1.07)	0.86 (0.80–0.92)	1.12 (1.04–1.22)	0.97 (0.86–1.10)	<0.0001
Thrombocytopenia	1.73 (1.56–1.92)	1.46 (1.30–1.63)	1.01 (0.83–1.23)	2.21 (1.84–2.64)	1.48 (1.10–1.99)	<0.0001
Neutropenia	1.33 (1.11–1.61)	1.40 (1.18–1.67)	1.25 (0.95–1.64)	2.47 (1.91–3.18)	1.63 (1.05–2.52)	0.0004
Monoclonal paraproteinaemia	2.01 (1.56–2.59)	1.80 (1.39–2.34)	1.45 (0.96–2.20)	3.43 (2.40–4.91)	1.51 (0.71–3.19)	0.0095
Cryoglobulinemia	4.58 (1.05–20.03)	6.36 (1.62–25.03)	10.14 (2.52–40.68)	10.23 (1.80–58.14)	6.64 (0.57–76.77)	0.6777

Condition	Adjusted odds ratio (95%CI)*					Heterogeneity p-value
	CLL/SLL	DLBCL	FL	MZL	TCL	
CARDIOVASCULAR DISORDERS						
Benign/unspecified hypertension	0.83 (0.80–0.86)	0.93 (0.89–0.96)	0.87 (0.83–0.92)	0.80 (0.74–0.86)	0.85 (0.77–0.94)	0.0001
Systolic heart failure	0.89 (0.85–0.94)	0.92 (0.88–0.96)	0.80 (0.74–0.86)	0.70 (0.63–0.78)	0.77 (0.67–0.88)	<0.0001
Abdominal aneurysm	0.67 (0.58–0.77)	1.05 (0.93–1.19)	0.84 (0.68–1.02)	0.92 (0.71–1.20)	0.88 (0.61–1.26)	0.0001
ENDOCRINE/METABOLIC CONDITIONS						
Diabetes mellitus	0.96 (0.92–1.00)	1.09 (1.05–1.13)	0.89 (0.84–0.94)	0.77 (0.71–0.83)	0.93 (0.84–1.04)	<0.0001
Hyperlipidemia NOS	0.86 (0.83–0.90)	0.96 (0.92–0.99)	0.97 (0.92–1.01)	0.85 (0.80–0.91)	0.89 (0.81–0.98)	0.0000
Testicular hypofunction	0.89 (0.73–1.09)	0.90 (0.73–1.11)	1.10 (0.84–1.45)	1.58 (1.15–2.17)	0.76 (0.43–1.35)	0.0175
NEUROLOGICAL/PSYCHIATRIC DISORDERS						
Dementia, senile or other causes	0.61 (0.56–0.66)	0.44 (0.40–0.48)	0.40 (0.35–0.47)	0.41 (0.34–0.49)	0.48 (0.37–0.62)	<0.0001
Arteriosclerotic dementia	0.54 (0.44–0.67)	0.35 (0.27–0.45)	0.32 (0.21–0.49)	0.33 (0.19–0.55)	0.50 (0.26–0.93)	0.0312
Parkinson's disease	0.74 (0.65–0.85)	0.67 (0.59–0.77)	0.71 (0.58–0.87)	0.56 (0.41–0.75)	0.62 (0.42–0.91)	0.3946
Stroke, including TIA	0.77 (0.74–0.81)	0.86 (0.83–0.91)	0.82 (0.76–0.88)	0.74 (0.67–0.81)	0.79 (0.69–0.90)	0.0032
Depression NOS	0.82 (0.76–0.87)	0.74 (0.69–0.80)	0.70 (0.63–0.78)	0.71 (0.62–0.81)	0.85 (0.70–1.03)	0.0519
Psychosis NOS	0.64 (0.56–0.74)	0.47 (0.40–0.56)	0.38 (0.28–0.50)	0.37 (0.26–0.53)	0.43 (0.27–0.70)	<0.0001
Nonpsychotic mental disorder NOS	0.55 (0.41–0.73)	0.23 (0.15–0.35)	0.29 (0.16–0.54)	0.16 (0.05–0.50)	0.67 (0.32–1.42)	0.0021
Alcoholism	0.79 (0.68–0.92)	0.59 (0.50–0.71)	0.60 (0.46–0.78)	0.81 (0.59–1.11)	0.98 (0.68–1.42)	0.0235
SKIN DISEASES						
Non-melanoma skin cancer	1.19 (1.13–1.25)	1.20 (1.15–1.26)	1.25 (1.17–1.34)	1.24 (1.13–1.35)	1.55 (1.37–1.76)	0.0018
Actinic keratosis	1.12 (1.07–1.17)	1.14 (1.09–1.19)	1.20 (1.13–1.28)	1.17 (1.08–1.28)	1.25 (1.11–1.41)	0.1607
Atopic dermatitis	0.95 (0.79–1.13)	1.12 (0.96–1.30)	0.94 (0.73–1.21)	1.28 (0.97–1.70)	4.12 (3.25–5.22)	<0.0001
Contact dermatititis and other eczema	0.91 (0.86–0.96)	1.04 (0.99–1.10)	0.98 (0.91–1.06)	1.09 (0.98–1.20)	2.61 (2.33–2.93)	<0.0001
Dermatitis due to substances taken internally	0.93 (0.71–1.20)	0.94 (0.73–1.21)	0.75 (0.50–1.14)	1.29 (0.85–1.96)	4.05 (2.84–5.79)	<0.0001
Bullous skin diseases	0.97 (0.64–1.48)	1.14 (0.78–1.67)	1.10 (0.61–1.97)	0.81 (0.33–1.98)	3.43 (1.82–6.45)	0.0113
Discoid lupus	0.35 (0.16–0.79)	1.93 (1.35–2.76)	0.55 (0.23–1.35)	1.19 (0.52–2.70)	4.00 (2.04–7.85)	<0.0001
Psoriasis and similar disorders	0.82 (0.71–0.95)	1.25 (1.11–1.40)	0.98 (0.82–1.19)	1.34 (1.07–1.67)	3.72 (3.05–4.53)	<0.0001
Seborrheic keratosis	1.01 (0.95–1.07)	1.00 (0.95–1.06)	1.18 (1.09–1.28)	1.24 (1.12–1.37)	1.25 (1.08–1.46)	<0.0001
Folliculitis and related conditions	1.00 (0.78–1.29)	1.10 (0.85–1.41)	0.54 (0.32–0.89)	1.25 (0.79–1.97)	2.65 (1.73–4.07)	<0.0001
Asteatosis	0.86 (0.72–1.02)	0.98 (0.84–1.15)	1.00 (0.79–1.26)	1.07 (0.80–1.43)	2.20 (1.62–2.97)	<0.0001
Urticaria	0.96 (0.78–1.18)	1.01 (0.84–1.22)	1.15 (0.89–1.50)	0.97 (0.67–1.40)	2.42 (1.71–3.41)	<0.0001

Condition	Adjusted odds ratio (95%CI)[*]					Heterogeneity p-value
	CLL/SLL	DLBCL	FL	MZL	TCL	
MISCELLANEOUS CONDITIONS						
Benign prostate hyperplasia	1.04 (0.99–1.10)	0.97 (0.91–1.03)	1.10 (1.02–1.20)	1.30 (1.16–1.44)	0.95 (0.82–1.10)	<0.0001
Gastric ulcer	0.92 (0.81–1.04)	0.88 (0.78–1.00)	0.88 (0.72–1.06)	1.55 (1.28–1.88)	0.84 (0.59–1.20)	<0.0001
Decubitus ulcer	0.76 (0.66–0.88)	0.57 (0.48–0.67)	0.56 (0.43–0.72)	0.48 (0.34–0.68)	0.64 (0.41–0.98)	0.0183
Specified anomalies of spinal cord (myelodysplasia)	3.20 (2.06–4.98)	1.69 (0.97–2.94)	0.57 (0.14–2.35)	2.23 (0.89–5.61)	1.93 (0.48–7.81)	0.0912
Asphyxia	0.90 (0.78–1.04)	0.79 (0.68–0.92)	0.52 (0.40–0.69)	0.60 (0.43–0.83)	1.00 (0.68–1.46)	0.0020
Hip fracture	0.76 (0.66–0.86)	0.62 (0.54–0.71)	0.50 (0.39–0.64)	0.71 (0.55–0.92)	0.61 (0.40–0.95)	0.0286

Abbreviations: HIV human immunodeficiency virus, HCV hepatitis C virus, NOS not otherwise specified, TIA transient ischemic attack

[*]
Odds ratios are adjusted for sex, age and calendar year of case/control selection, race, and number of provider claims per year. Results are not shown for models that failed to converge. Odds ratios that are shaded were identified as SELECT3 conditions for the indicated NHL subtype, and underlining indicates statistical significance.

Table 4

Replication of selected associations in additional populations

Medical condition	Population	Number of people with medical condition in sample	Relative risk (95% confidence interval)*				
			CLL/SLL	DLBCL	FL	MZL	TCL
Hypertension							
	MedWAS	--	0.83 (0.80–0.86)	0.93 (0.89–0.96)	0.87 (0.83–0.92)	0.80 (0.74–0.86)	0.85 (0.77–0.94)
	NIH-AARP	129,350	0.96 (0.82–1.12)	1.01 (0.85–1.21)	1.02 (0.83–1.26)	0.63 (0.45–0.88)	0.77 (0.53–1.13)
	PLCO	50,400	1.18 (0.99–1.40)	1.02 (0.82–1.27)	0.87 (0.66–1.16)	--†	1.08 (0.65–1.80)
Diabetes mellitus							
	MedWAS	--	0.96 (0.92–1.00)	1.09 (1.05–1.13)	0.89 (0.84–0.94)	0.77 (0.71–0.83)	0.93 (0.84–1.04)
	NIH-AARP	50,748	1.10 (0.91–1.33)	1.17 (0.95–1.45)	0.94 (0.72–1.24)	0.73 (0.45–1.19)	1.85 (1.25–2.73)
	PLCO	11,364	1.28 (0.95–1.73)	1.07 (0.72–1.60)	1.11 (0.67–1.82)	--†	1.49 (0.68–3.28)
Stroke							
	MedWAS	--	0.77 (0.74–0.81)	0.86 (0.83–0.91)	0.82 (0.76–0.88)	0.74 (0.67–0.81)	0.79 (0.69–0.90)
	NIH-AARP	11,927	0.98 (0.67–1.44)	1.20 (0.80–1.80)	1.04 (0.61–1.76)	0.54 (0.17–1.68)	0.97 (0.36–2.62)
	PLCO	2421	1.30 (0.78–2.17)	1.62 (0.91–2.88)	1.47 (0.69–3.12)	--†	2.14 (0.67–6.83)
Chronic bronchitis							
	MedWAS	--	0.82 (0.77–0.89)	0.78 (0.73–0.84)	0.79 (0.71–0.88)	0.95 (0.83–1.08)	0.73 (0.60–0.89)
	PLCO	7071	0.42 (0.22–0.78)	0.72 (0.39–1.31)	1.31 (0.75–2.29)	--†	1.10 (0.35–3.52)

Abbreviations:: NHL non-Hodgkin lymphoma, CLL/SLL chronic lymphocytic leukemia/small lymphocytic lymphoma, DLBCL diffuse large B-cell lymphoma, MZL marginal zone lymphoma, TCL T-cell lymphoma, NIH-AARP National Institutes of Health-AARP Diet and Health Study, PLCO Prostate Lung Colorectal and Ovarian Cancer Screening Trial.

*
Relative risks from the MedWAS study are highlighted in gray if they were identified in the SELECT3 medical conditions in the MedWAS analyses (see Table 2). Relative risks are underlined if they are statistically significant (p<0.05). Relative risks were calculated in the MedWAS analyses as odds ratios and in the NIH-AARP and PLCO analyses as hazard ratios.

†
There were no cases of marginal zone lymphoma in the PLCO cohort.