



HHS Public Access

Author manuscript

Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2020 November 01.

Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2020 May ; 29(5): 936–941. doi:10.1158/1055-9965.EPI-19-1504.

Assessing cancer treatment information using Medicare and hospital discharge data among women with non-Hodgkin lymphoma in a Los Angeles County case-control study

Charlie Zhong¹, Petra Seibold^{2,3}, Chun R. Chao⁴, Wendy Cozen⁵, Joo Y. Song⁶, Dennis Weisenburger⁶, Leslie Bernstein², Sophia S. Wang¹

¹Division of Health Analytics, Department of Computational and Quantitative Medicine, City of Hope and the Beckman Research Institute, Duarte, CA, USA

²Department of Population Sciences, City of Hope and the Beckman Research Institute, Duarte, CA, USA

³Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany

⁴Department of Research & Evaluation, Kaiser Permanente Southern California, Pasadena, CA, USA

⁵Department of Preventive Medicine, University of Southern California, Los Angeles, CA, USA

⁶Department of Pathology, City of Hope National Medical Center, Duarte, CA, USA

Abstract

Background: We assessed the ability to supplement existing epidemiological/etiologic studies with data on treatment and clinical outcomes by linking to publicly available cancer registry and administrative databases.

Methods: Medical records were retrieved and abstracted for cases enrolled in a Los Angeles County case-control study of non-Hodgkin Lymphoma(NHL). Cases were linked to the Los Angeles County cancer registry(CSP), the California state hospitalization discharge database(OSHPD), and the Surveillance, Epidemiology and End Results Registry(SEER)-Medicare database. We assessed sensitivity, specificity, and positive predictive value(PPV) of cancer treatment in linked databases, compared to medical record abstraction.

Results: We successfully retrieved medical records for 918 of 1,004 participating NHL cases and abstracted treatment for 698. We linked 59% of cases(96% of cases >65 years old) to SEER-Medicare and 96% to OSHPD. Chemotherapy was the most common treatment and best captured, with the highest sensitivity in SEER-Medicare(80%) and CSP(74%); combining all three data sources together increased sensitivity(92%), at reduced specificity(56%). Sensitivity for radiotherapy was moderate; 77% with aggregated data. Sensitivity of BMT was low in the

To whom correspondence should be addressed: Charlie Zhong, Ph.D., Division of Health Analytics, Department of Computational & Quantitative Medicine, Beckman Research Institute of the City of Hope, 1500 East Duarte Road, Duarte, CA 91010, Phone: (626) 218-2170, Fax: (626) 471-7308 czhong@coh.org.

Conflict of interest: The authors declare no potential conflicts of interest.

CSP(42%), but high for the administrative databases, especially OSHPD(98%). Sensitivity for surgery reached 83% when considering all three datasets in aggregate, but PPV was 60%. In general, sensitivity and PPV for chronic lymphocytic leukemia/small lymphocytic lymphoma(CLL/SLL) were low.

Conclusions: Chemotherapy was accurately captured by all data sources. Hospitalization data yielded the highest performance values for BMTs. Performance measures for radiotherapy and surgery were moderate.

Impact: Various administrative databases can supplement epidemiological studies depending on treatment type and NHL subtype of interest.

Keywords

non-Hodgkin lymphoma; administrative databases; case-control study; cancer registry; cancer treatment

Introduction

The American Cancer Society (ACS) estimates there are over 750,000 non-Hodgkin lymphoma (NHL) survivors in the United States in 2019, and projects this number to increase to over 1 million by 2030(1). Long-term health-related morbidities, such as cardiovascular complications and secondary malignancies will also increase with increasing survivorship(2–5). Continued research among NHL survivors is thus important for improving long-term outcomes. In addition to clinical studies of patient populations, existing epidemiological studies which have largely focused on etiologic research, could methodologically and efficiently be leveraged to complement on-going NHL outcome/survivorship research(6). Because these epidemiological studies have traditionally collected pre-diagnostic information, they provide a key component for stratifying outcomes that clinically-based survivorship studies may not have. Several methodological studies assessing whether one can ascertain treatment data for breast, colorectal, lung and prostate cancer patients participating in epidemiologic studies have been conducted(7–10), but fewer studies have assessed whether treatment patterns and clinical data can be ascertained for NHL patient participants(11,12). Here, we evaluate the use of readily available cancer registry and administrative databases to ascertain cancer treatment data for a large group of women diagnosed with incident B-cell NHL between 2004 and 2008 in Los Angeles County, California(13).

Materials and Methods

Study population

Our study population comprised 1,004 women with B-cell NHL diagnosed between May 2004 and March 2008 who were identified by the population-based Los Angeles County Cancer Surveillance Program (CSP)(13) and consented to participate in a case-control study of incident NHL which included the following eligible NHL diagnoses: diffuse large B-cell lymphoma (DLBCL: 9678, 9679, 9680, 9684), follicular lymphoma (FL: 9690, 9691, 9695, 9698), mantle cell lymphoma (MCL: 9673), marginal zone lymphoma (MZL: 9699), chronic

lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL: 9670, 9823), Burkitt lymphoma (BL: ICD-O-3 codes 9687, 9826), and other B-cell lymphomas (Other, including not otherwise specified (NOS): 9590, 9591, 9596, 9671, 9675, 9727, 9728, 9833, 9835, 9836, 9761). Informed written consent was obtained according to the Declaration of Helsinki and this study is approved by the Institutional Review Boards of the City of Hope, University of Southern California, Kaiser Permanente Southern California, and Committee for the Protection of Human Subjects of the California Health and Human Services Agency in accordance with assurances filed and approved by the U.S. Department of Health and Human Services.

Data sources

We collected NHL treatment data from four overlapping data sources: (i) medical record retrieval and abstraction; (ii) Cancer Surveillance Program (CSP); (iii) Surveillance, Epidemiology and End Results Registry (SEER)-Medicare; and (iv) California state hospitalization inpatient discharge database (OSHPD). Linkages to the hospitalization and SEER-Medicare databases were conducted by the California Cancer Registry and SEER-Medicare, respectively, using zip code at diagnosis, social security number, date of birth and sex(14).

- i. Medical records retrieval was conducted from July 2014 to December 2016 at all CSP identified hospitals. Medical record abstraction was conducted by three trained reviewers using a standardized abstraction form. Data abstracted from medical records were considered the gold standard and to which the three other data sources were compared for completeness. We abstracted information on diagnosis, relapse, outcome, chemotherapy regimens received, surgical treatment and site, radiotherapy field and dose, and bone marrow transplant.
- ii. Cancer Surveillance Program (CSP) is the population-based cancer registry for Los Angeles County with near complete incidence data for the area since 1972(15). The registry collects initial treatment data and is also part of the Surveillance, Epidemiology, and End Results (SEER) program. All 1,004 NHL cases were identified through the CSP and have data on date of their initial course of treatment including; chemo/immunotherapy, radiotherapy, transplant, and surgery. The cancer registry also links to death indices to obtain mortality data. In addition, we were able to obtain additional free text information on chemo/immunotherapy regimen.
- iii. Hospital Discharge Data of the Office of Statewide Health Planning and Development (OSHPD) in California. Non-federal hospitals are required to report semiannually 18 data elements per patient to OSHPD(16). At the time of the analysis, OSHPD data were available from 1999–2014. Of the 1,004 NHL cases, 959 were linked to a discharge record within a year of their NHL diagnosis.
- iv. SEER-Medicare: Medicare is the primary health care insurer for the vast majority of the US population that is 65 years or older and some individuals who are disabled or have end stage renal disease. Basic coverage (Part A) includes

inpatient care (MEDPAR file). More than 90% were also enrolled in Part B covering physician services, outpatient care, and durable medical equipment in 2015. About 70% of all enrollees had additional Part D outpatient prescription drug coverage (implemented in 2006) in 2014(17). As part of a collaborative effort between the National Cancer Institute's SEER registry program, and the Centers for Medicare and Medicaid Services (CMS); 95 percent of patients aged 65 or older identified in the SEER registry are successfully matched to their Medicare enrollment file(18). Medicare data for the NHL study were available for the years 1991–2014 (2007–2014 for Part D data).

Cancer treatment abstraction from administrative databases

Cancer treatments abstracted were: (1) chemotherapy/immunotherapy/targeted therapy, (2) radiotherapy, and (3) bone marrow/stem cell transplantation (BMT). We also abstracted for tumor surgery to determine whether initial biopsies could be delineated from tumor surgeries. ICD-9 diagnosis, procedure codes, diagnosis related groups (DRG) codes, and Healthcare Common Procedure Coding System (HCPCS)/Current Procedural Terminology (CPT) codes were utilized to identify these treatments (Supplementary Table S1). For the purposes of this analysis, we compared first course of treatment identified in each data source that was documented after the date of the patient's confirmed diagnosis. All databases linked exceeded the necessary follow-up period (through 2014) that this would encompass initial treatments, as NHL diagnoses were from 2004–2008.

Statistical methods

We evaluated characteristics and the availability of specific treatment information for NHL overall and the three most common NHL subtypes for which there was sufficient numbers to evaluate (DLBCL, FL, and CLL/SLL). Each administrative database was assessed for sensitivity, specificity, and positive predictive value (PPV) by treatment type for all NHL and by NHL subtype, with data from medical record abstraction considered as the "gold standard". Sensitivity was calculated as the number of cases identified by both the medical record and administrative database as having received treatment divided by the total number of cases identified by medical records as having received that treatment. Specificity was calculated as the number of cases without the treatment type abstracted by medical records and without record of that treatment in the administrative database divided by the total number of cases without that treatment according to medical record abstraction. PPV was calculated as the number of cases identified by both the medical record and administrative database as having received treatment divided by the total number of cases with that treatment according to the administrative database. To ensure comparability of data, we also conducted the following sensitivity analyses and shown in supplemental data tables. First, analyses of treatment data in SEER-Medicare were restricted to eligible cases 65 years and older. Second, we also conducted sensitivity analysis where we restricted the definition of the first reported/course of treatment to be within a year of diagnosis. We further evaluated chemotherapy regimen text field data available from the CSP in DLBCL and FL. All analyses were conducted using SAS 9.4 (SAS Institute Inc., Cary, NC).

Results

Characteristics of the study population

Of the 1,004 cases, 965 consented to medical record retrieval. Of those, we requested and successfully received medical records for 918 (95%) cases. For the purposes of this analysis, we include abstracted treatment information for 698 (76%) cases for which treatment information could be confidently ascertained (Supplementary Table S2). We note that the availability of treatment information differed by NHL subtype, with the most complete information ascertained for DLBCL (91%), followed by FL (75%), and was lowest for CLL/SLL (55%). OSHPD hospital discharge records were available for 959 (96%) cases (Supplementary Figure S1). SEER-Medicare records were available for a total of 596 (59%) cases, of which 438 were not enrolled in a health maintenance organization (HMO) and therefore had claims data. Case characteristics by administrative data source are presented in Table 1. As expected, cases with Medicare records were older (median age: 69 years) compared to the general study population (median age: 61 years). By NHL subtype, there were proportionally more CLL/SLL cases in the Medicare data, likely due to their higher age of diagnosis. Cases in SEER-Medicare also had lower annual household incomes, another attribute of the Medicare population which would not have records of claims submitted to alternate health insurance providers such as through their employer or HMO.

Chemotherapy

Compared to medical record abstraction, sensitivity (of correctly identifying all patients who had chemotherapy) was highest in the Medicare subgroup (80%), and relatively high in the CSP (74%); combining all three data sources together increased sensitivity to 92% (Table 2). The specificity (correctly identifying which patients did not have chemotherapy) was high in the CSP (99%) for all NHL, with the exception of DLBCL (67%), but we note that this measure was based on only a handful of DLBCL patients (n=6) who did not in fact have chemotherapy based on medical records (Supplementary Table S2). Specificity for OSHPD data was 72% overall and 100% for DLBCL. Although leveraging data from all three data sources to identify chemotherapy among cases increased sensitivity, it reduced specificity (56%). Finally, PPV was high overall for all the data sources (91–100%) for NHL, DLBCL, and FL. We note that for CLL/SLL, sensitivity of CSP and OSHPD data was low (34% and 57%, respectively), and high in the Medicare subset (94%). Although the combination of all three data sources increased overall sensitivity to 81%, this appears to be attributed to the more complete data ascertained from the Medicare subset; sensitivity among the non-Medicare subset of participants remained low. Overall, sensitivity, specificity, and PPV were generally similar in CSP and OSHPD when restricted to those 65 and older at diagnosis, but individual values within the Medicare population subset improved (Supplementary Table S3). In summary, patients identified by the CSP, OSHPD, and Medicare as having chemotherapy could be confirmed as having chemotherapy in medical records, with the exception of CLL/SLL. However, to identify all patients indicated by medical records as having chemotherapy, there were clear benefits to combining all three data sources to achieve higher sensitivity, although at a cost of reduced specificity.

Comparison of chemotherapy regimens recorded

We further evaluated the ability to identify specific chemotherapy regimens in the CSP data for DLBCL, in which the standard of care has included the R-CHOP (rituximab in combination with cyclophosphamide, doxorubicin, vincristine, and prednisone) regimen since the early 2000s. Of 253 DLBCL cases abstracted, 52 did not have chemotherapy regimen in CSP data. Among the 208 cases that received CHOP or R-CHOP according to medical record abstraction, 150 (72%) were identified in the CSP; we note that 16 of those cases were recorded as having received only CHOP by the CSP when medical records specified treatment for R-CHOP (Supplementary Table S4). For the 149 FL cases that received chemotherapy, 49 were missing regimen information in the CSP data. Medical records identified 69 cases as receiving CHOP or R-CHOP, and 46 (67%) of those were also identified in the CSP (Supplementary Table S5). Only 8 (38%) of the 21 cases that received R-CVP/CVP were correctly identified in the CSP regimen text field.

Radiotherapy

Overall, sensitivity for identifying patients who underwent radiotherapy using the three data sources was low (11–67%), though using the data sources in aggregate did increase sensitivity to 71–77% for overall, DLBCL, and follicular lymphoma. However, specificity and PPV were high, largely driven by the larger numbers of participants who did not have radiotherapy. Values for CLL/SLL were unstable due to the relatively small number of participants who had radiotherapy.

Bone marrow transplant (BMT)

The sensitivity of identifying patients who underwent bone marrow transplant was low in the CSP (42% overall and 25% for DLBCL), but high for the administrative databases, especially OSHPD (98% overall, 96% for DLBCL). The aggregated data yielded 98% sensitivity overall, and 96% for DLBCL; specificity also remained high (98%). PPV was moderately high for OSHPD and Medicare, but we note that for the BMTs that the CSP data did capture, PPV was 100%.

Surgery

Although surgery is not a common treatment for lymphoma, we included medical abstraction for surgery to assess whether initial biopsies could be delineated from tumor surgeries in the various databases, as can be delineated based on medical record abstraction. The sensitivity of identifying patients who had undergone surgery in the three data sources varied from 22–73% overall and reached 83% when considering all three datasets in aggregate. Again, the increase in sensitivity resulted in reduced specificity. By NHL subtype, sensitivity was highest in the CSP dataset for both DLBCL (84%) and FL (76%), but low (31%) for CLL/SLL. Interestingly, aggregating the three datasets did not yield substantive benefits in sensitivity for DLBCL and FL, beyond what was already achieved by the CSP data. PPV was 63–68% overall and for DLBCL and FL, indicating that a third of participants who were indicated by the data sources to having had surgery in fact did not have surgery. Sensitivity for CLL/SLL was poor and reached 54% with the aggregated data; PPV was highest in the OSHPD/hospitalization data (75%).

Discussion

In this comparison of NHL treatment data by multiple data sources, we found that the cancer treatments captured vary by the data source and by the NHL subtype, but that the aggregated data sources improved performance measures in general. Specifically, sensitivity of identifying patients who had chemotherapy was high in the CSP/cancer registry data and Medicare for overall NHL and for DLBCL and FL. Although aggregating the data improved sensitivity further, this resulted in a reduced specificity. Nevertheless, PPV and thus confidence that those identified by the data sources as having had chemotherapy in fact had chemotherapy, was high.

Our further evaluation of specific chemotherapy regimens further demonstrated good agreement between chemotherapy regimens reported by the cancer registry to medical record abstraction. We were unable to identify specific regimen or agents in SEER-Medicare with high accuracy. While the administrative payment data in SEER-Medicare does include charges related to specific chemotherapeutic agents and, uniquely, dose, these data only provide the agents used in an inpatient setting, making identification of multi-agent regimens difficult. Any prescriptions, such as prednisone, one of the agents in the R-CHOP regimen, would only be captured for the subset of cases with Part D coverage starting in 2006; and because diagnoses are not associated with these prescriptions, these data could falsely identify cases as having received treatment. When we excluded Part D coverage, we were able to increase specificity for chemotherapy, but this came at the cost of greatly reduced sensitivity. Of the 37 DLBCL patients identified in SEER-Medicare, and not part of a managed care organization, that received R-CHOP according to medical records, we were only able to identify 12 in SEER-Medicare with payment codes for all 5 agents.

Overall, the performance values for identifying or confirming patients who had radiotherapy were lower than chemotherapy, but better than surgery. Due to the typical outpatient nature of radiotherapy, it was not captured well in OSHPD. In contrast, BMT performed much better in OSHPD/hospitalization data due to the inpatient nature of the procedure. The ascertainment of surgery proved more challenging as most surgical procedures for NHL are typically done for diagnostic and not treatment purposes. Surgical resection is more common for gastric NHL and if disease is localized to certain areas such as the spleen(19). Sensitivity with the three data sources together was ~80%. Again, this increase in sensitivity came at a cost of reduced specificity. PPV was modest (and lower than that for chemotherapy), reflecting the poor distinction between initial biopsies versus the excisional treatment, as there are no specific surgery codes related to NHL surgery (in contrast to mastectomy, for example).

While performance values were largely similar between NHL overall, DLBCL, and FL, they were generally low for CLL/SLL. Sensitivity of CSP and OSHPD data for chemotherapy and surgery were low, reflecting the common watch-and-wait strategy for this NHL subtype which was confirmed in medical record abstraction. Of the three data sources, Medicare data yielded the highest sensitivity for capturing chemotherapy, surgery, and radiotherapy among CLL/SLL, and would seem a likely resource for CLL/SLL research restricted to older adults.

Compared to the time and effort required for medical record retrieval and abstraction, there are clear benefits for larger-scale data linkage efforts for retrospective analyses. Although we were able to obtain medical records for a majority of our cases, only 76% were considered sufficient for confidently ascertaining treatment information. Depending on the subtype and treatment of interest, linkage to different data sources may provide efficient ways to obtain treatment data. Examples of specific research questions could entail: (i) research on CLL/SLL within Medicare data; (ii) leveraging hospitalization data for BMT research; or (iii) utilizing all three data sources for ascertainment of chemotherapy and cancer registry data for chemotherapy regimens, particularly for DLBCL research.

Limitations of administrative databases include that both SEER-Medicare and OSHPD linkages are based on social security number, zip code, date of birth, and sex. Therefore, there may be no matches for a small proportion of cases due to missing or incorrect social security number (up to 10% for OSHPD) (M. Allen/CCR, personal communication). Some of the linkages, such as OSHPD, were required to be conducted by an outside entity; we were therefore not able to distinguish between “no hospitalization” and “linkage failed”. Age remains a well-known limitation for linkage to SEER-Medicare records. 59% of our cases were younger than 65 years at their NHL diagnosis and most of them were therefore not yet eligible for Medicare. Potential socioeconomic and racial biases may also exist with Medicare-restricted populations(20). In addition to being 65 years or older to receive Medicare coverage, individuals are also required to have at least five years residency in the US and must have been working for at least ten years in Medicare-covered employment in order to be eligible. And of course, cases who are members of an HMO will not have records in the Medicare database as their claims would be processed by the HMO. Kaiser Permanente Southern California is a large HMO provider that provides service to a third of the residents in the Southern California region, including Los Angeles County(21). In our study, between 24% and 31% per annum were enrolled in managed care. As our study focused on California residents and leveraged the California hospitalization discharge database (OSHPD), cases who received cancer treatment in another state may be missed by our approach. Finally, for certain groups such as former military members diagnosed and treated in veteran’s administration (VA) hospitals, cancer information will likely be missing in the CSP, Medicare, or OSHPD.

In summary, leveraging cancer registry and readily available administrative data sources provided informative data for NHL treatments, especially chemotherapy. Depending on the target population and scientific research questions of interest (e.g., older patients, BMT survivors, specific NHL subtypes), different approaches and linkages can be taken, drawing upon the strengths of each complementary data source to successfully accomplish the intended research aims.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the National Cancer Institute under grants, R01 CA166219 to SS Wang, and R01 CA108634, P01 CA017054, and P30 CA033572 to L Bernstein.

This study used the linked SEER-Medicare database. The interpretation and reporting of these data are the sole responsibility of the authors. The authors acknowledge the efforts of the National Cancer Institute; the Office of Research, Development and Information, CMS; Information Management Services (IMS), Inc.; and the Surveillance, Epidemiology, and End Results (SEER) Program tumor registries in the creation of the SEER-Medicare database.

The collection of cancer incidence data used in this study was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract HHSN261201000140C awarded to the Cancer Prevention Institute of California, contract HHSN261201000035C awarded to the University of Southern California, and contract HHSN261201000034C awarded to the Public Health Institute; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement # U58DP003862-01 awarded to the California Department of Public Health. The ideas and opinions expressed herein are those of the author(s) and endorsement by the State of California Department of Public Health, the National Cancer Institute, and the Centers for Disease Control and Prevention or their Contractors and Subcontractors is not intended nor should be inferred. The authors acknowledge the efforts of the National Cancer Institute; the Office of Research, Development and Information, CMS; Information Management Services (IMS), Inc.; and the Surveillance, Epidemiology, and End Results (SEER) Program tumor registries in the creation of the SEER-Medicare database.

We gratefully thank Susan Gundell, Kathy Lane, Cynthia Quince and Teri Terrusa for their efforts consenting and interviewing participants and Kimberly Cannavale, Michelle Dich, Sunhea Sylvia Kim, Diana Reyes, and Olivia Sattayapiwat for their efforts abstracting the medical records.

References

1. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. *CA: a cancer journal for clinicians* 2019 doi 10.3322/caac.21565.
2. Pirani M, Marcheselli R, Marcheselli L, Bari A, Federico M, Sacchi S. Risk for second malignancies in non-Hodgkin's lymphoma survivors: a meta-analysis. *Ann Oncol* 2011;22(8):1845–58 doi 10.1093/annonc/mdq697. [PubMed: 21310758]
3. Ruiz-Soto R, Sergent G, Gisselbrecht C, Larghero J, Ertault M, Hennequin C, et al. Estimating late adverse events using competing risks after autologous stem-cell transplantation in aggressive non-Hodgkin lymphoma patients. *Cancer* 2005;104(12):2735–42 doi 10.1002/cncr.21492. [PubMed: 16284986]
4. Blaes A, Konety S, Hurley P. Cardiovascular Complications of Hematopoietic Stem Cell Transplantation. *Current treatment options in cardiovascular medicine* 2016;18(4):25 doi 10.1007/s11936-016-0447-9. [PubMed: 26909819]
5. Baker KS, Armenian S, Bhatia S. Long-term consequences of hematopoietic stem cell transplantation: current state of the science. *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation* 2010;16(1 Suppl):S90–6 doi 10.1016/j.bbmt.2009.09.017.
6. Berrington de Gonzalez A, Morton LM. Converting epidemiologic studies of cancer etiology to survivorship studies: approaches and challenges. *Cancer Epidemiol Biomarkers Prev* 2012;21(6):875–80 doi 10.1158/1055-9965.EPI-12-0131. [PubMed: 22426143]
7. Cooper GS, Virnig B, Klabunde CN, Schussler N, Freeman J, Warren JL. Use of SEER-Medicare data for measuring cancer surgery. *Medical care* 2002;40(8 Suppl):IV-43–8 doi 10.1097/01.MLR.0000020943.21850.F1.
8. Virnig BA, Warren JL, Cooper GS, Klabunde CN, Schussler N, Freeman J. Studying radiation therapy using SEER-Medicare-linked data. *Medical care* 2002;40(8):49–54 doi 10.1097/01.MLR.0000020940.90270.4D.
9. Warren JL, Harlan LC, Fahey A, Virnig BA, Freeman JL, Klabunde CN, et al. Utility of the SEER-Medicare data to identify chemotherapy use. *Medical care* 2002;40(8 Suppl):IV-55–61 doi 10.1097/01.MLR.0000020944.17670.D7.

10. Lund JL, Sturmer T, Harlan LC, Sanoff HK, Sandler RS, Brookhart MA, et al. Identifying specific chemotherapeutic agents in Medicare data: a validation study. *Medical care* 2013;51(5):e27–34 doi 10.1097/MLR.0b013e31823ab60f. [PubMed: 22080337]
11. Griffiths R, Gleeson M, Knopf K, Danese M. Racial differences in treatment and survival in older patients with diffuse large B-cell lymphoma (DLBCL). *BMC Cancer* 2010;10:625 doi 10.1186/1471-2407-10-625. [PubMed: 21073707]
12. Wang M, Burau KD, Fang S, Wang H, Du XL. Ethnic variations in diagnosis, treatment, socioeconomic status, and survival in a large population-based cohort of elderly patients with non-Hodgkin lymphoma. *Cancer* 2008;113(11):3231–41 doi 10.1002/ncr.23914. [PubMed: 18937267]
13. Wang SS, Luo J, Cozen W, Lu Y, Halley-Sullivan J, Voutsinas J, et al. Sun sensitivity, indoor tanning and B-cell non-Hodgkin lymphoma risk among Caucasian women in Los Angeles County. *Br J Haematol* 2017;177(1):153–6 doi 10.1111/bjh.14051. [PubMed: 27030403]
14. Potosky AL, Riley GF, Lubitz JD, Mentnech RM, Kessler LG. Potential for Cancer-Related Health-Services Research Using a Linked Medicare-Tumor Registry Database. *Medical care* 1993;31(8):732–48. [PubMed: 8336512]
15. Mack TM. Cancer surveillance program in Los Angeles County. National Cancer Institute monograph 1977;47:99–101. [PubMed: 613253]
16. California Office of Statewide Health Planning and Development. Reporting Requirements. 2014.
17. US Centers for Medicare & Medicaid Services. 2015 CMS Statistics. 2015.
18. 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. *Medical care* 2002;40(8 Suppl):IV-3–18 doi 10.1097/01.MLR.0000020942.47004.03.
19. Ansell SM. Non-Hodgkin Lymphoma: Diagnosis and Treatment. *Mayo Clinic proceedings* 2015;90(8):1152–63 doi 10.1016/j.mayocp.2015.04.025. [PubMed: 26250731]
20. Setiawan VW, Virnig BA, Porcel J, Henderson BE, Le Marchand L, Wilkens LR, et al. Linking data from the Multiethnic Cohort Study to Medicare data: linkage results and application to chronic disease research. *Am J Epidemiol* 2015;181(11):917–9 doi 10.1093/aje/kwv055. [PubMed: 25841869]
21. Permanente K 2017 10/02/2017 Fast Facts About Kaiser Permanente. <<https://share.kaiserpermanente.org/article/fast-facts-about-kaiser-permanente/>>. Accessed 2017 10/02/2017.

Table 1.

Characteristics of B-cell non-Hodgkin lymphoma (NHL) patients in cancer registry (California Surveillance Program, CSP) and Surveillance, Epidemiology, and End Results Program (SEER)-Medicare data.

Characteristics	Values	All CSP (N=1,004) N (%)	SEER-Medicare (N=596) N (%)	Not in SEER-Medicare (N=408)* N (%)
Coverage		100%	59%	41%
Age at diagnosis (years) ^α	Median (min-max)	61 (20-79)	69 (22-79)	48 (20-79)
	20-64 years	590 (59%)	200 (34%)	390 (96%)
	65-79 years	414 (41%)	396 (66%)	18 (4%)
NHL group ^α	CLL/SL/L	213 (21%)	159 (27%)	54 (13%)
	DLBCL	280 (28%)	145 (24%)	135 (33%)
	FL	245 (24%)	135 (23%)	110 (27%)
	MZL	130 (13%)	79 (13%)	51 (13%)
	NOS/Other	136 (14%)	78 (13%)	58 (14%)
Vital status ^α	Deceased through 2016	295 (29%)	218 (37%)	77 (19%)
NHL tumor stage ^α	Stage I-II	322 (32%)	166 (28%)	156 (38%)
	Stage III-IV	358 (36%)	216 (36%)	142 (35%)
Comorbidities ^β	diabetes, statin/baby aspirin use	376 (37%)	267 (50%)	109 (29%)
Race/Ethnicity ^β	Non-Hispanic White	627 (62%)	418 (70%)	209 (51%)
	Non-Hispanic Black	97 (10%)	58 (10%)	39 (10%)
	Hispanic	193 (19%)	89 (15%)	104 (25%)
	Asian/Other	87 (9%)	31 (5%)	56 (14%)
Education (years in school) ^β	9 years	44 (4%)	30 (5%)	14 (3%)
	10-12 years	253 (25%)	175 (29%)	78 (19%)
	13-14 years	338 (34%)	191 (32%)	147 (36%)
	15-16 years	365 (37%)	200 (34%)	169 (41%)

Characteristics	Values	All CSP (N=1,004) N (%)	SEER-Medicare (N=596) N (%)	Not in SEER-Medicare (N=408)* N (%)
Smoking ^β	6+ months	416 (41%)	259 (43%)	157 (39%)
Household income (annual) ^β	\$15,000-\$34,999	300 (30%)	200 (34%)	100 (25%)
	\$35,000-\$69,999	254 (25%)	152 (26%)	102 (25%)
	\$70,000+	351 (35%)	174 (29%)	177 (43%)
	Missing	99 (10%)	70 (12%)	29 (7%)

CLL/SLL: Chronic lymphocytic leukemia/small lymphocytic lymphoma. DLBCL: Diffuse large B-cell lymphoma. FL: Follicular lymphoma. MCL: Mantle cell lymphoma. MZL: Marginal zone lymphoma. NOS: Not otherwise specified. CSP: Cancer Surveillance Program Los Angeles. SEER: Surveillance, Epidemiology, and End Results Program.

*Numbers smaller than 11 not reported for patient confidentiality.

^αObtained from Cancer Surveillance Program

^βObtained from questionnaire

Table 2.

Sensitivity, specificity, and positive predictive value (PPV) of treatment (chemotherapy, surgery, radiotherapy, bone marrow transplant (BMT)) data abstracted from cancer registry (California Surveillance Program), OHSPD/hospitalization discharge, and SEER-Medicare data, compared to treatment abstracted from medical records among B-cell NHL cases diagnosed in Los Angeles County (2004–08).

Treatment	Overall			DLBCL			FL			CLL/SLL						
	CSP N=695	OSHPD N=665	Medicare N=401	Combined N=698	CSP N=243	OSHPD N=235	Medicare N=133	Combined N=243	CSP N=167	OSHPD N=162	Medicare N=97	Combined N=167	CSP N=105	OSHPD N=98	Medicare N=70	Combined N=105
Chemotherapy																
Sensitivity (%)	74	65	80	92	88	67	76	97	65	59	70	90	34	57	94	81
Specificity (%)	99	72	51	56	67	100	*	67	100	68	90	70	100	86	52	63
PPV (%)	100	94	91	93	100	100	100	100	100	92	98	95	100	91	80	83
Surgery																
Sensitivity (%)	73	22	52	83	84	41	63	83	76	10	52	81	31	12	50	54
Specificity (%)	76	90	74	61	75	84	63	61	45	89	77	36	92	99	87	86
PPV (%)	68	60	61	60	63	58	49	51	67	56	80	65	57	75	62	56
Radiotherapy																
Sensitivity (%)	67	11	41	77	61	12	32	71	61	3	40	75	50	*	100	100
Specificity (%)	99	97	89	91	100	96	86	89	99	98	87	92	99	*	93	96
PPV (%)	95	57	49	75	100	62	47	77	96	33	46	75	50	*	25	33
BMT																
Sensitivity (%)	42	98	87	98	25	96	88	96	*	*	*	*	*	*	*	*
Specificity (%)	100	98	98	98	100	98	99	98	*	*	*	*	*	*	*	*
PPV (%)	100	87	72	83	100	85	88	85	*	*	*	*	*	*	*	*

* insufficient number of cases receiving treatment to calculate

CLL/SLL: Chronic lymphocytic leukemia/small lymphocytic lymphoma. DLBCL: Diffuse large B-cell lymphoma. FL: Follicular lymphoma.