# Validation of an electronic algorithm for Hodgkin and non-Hodgkin Lymphoma in ICD-10-CM

Mara M. Epstein[1,2], Sarah K. Dutcher[3], Judith C, Maro[4], Cassandra Saphirak[1,2], Sandra DeLuccia[4], Muthalagu Ramanathan[5], Tejaswini Dhawale[6], Sonali Harchandani[5], Christopher Delude[2], Laura Hou[4], Autumn Gertz[4], Nina DiNunzio[4], Cheryl N. McMahill-Walraven[7], Mano S. Selvan[8], Justin Vigeant[4], David V. Cole[4], Kira Leishear[3], Jerry H. Gurwitz[1,2], Susan Andrade[1,2], Noelle M. Cocoros[4]

[1]Division of Geriatric Medicine, Department of Medicine, University of Massachusetts Medical School; [2]The Meyers Primary Care Institute, a joint endeavor of the University of Massachusetts Medical School, Reliant Medical Group, and Fallon Health; [3]Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, Food and Drug Administration; [4]Department of Population Medicine, Harvard Pilgrim Health Care Institute, Harvard Medical School; [5]Division of Hematology and Oncology, Department of Medicine, UMass Memorial Medical Center; [6]Division of Medicine, Massachusetts General Hospital; [7]Aetna, a CVS Health company; [8]Humana Healthcare Research, Inc.

## OBJECTIVE

- Develop and validate an ICD-10-CM-based algorithm to identify lymphoma with a positive predictive value (PPV) of at least 80% (±10%) to support studies of drug safety

## BACKGROUND

- >77,000 new non-Hodgkin (NHL) and 8,400 Hodgkin lymphoma (HL) cases diagnosed in the US each year
- More than 60 histologic subtypes of lymphoma
- Lymphoma is generally diagnosed in the outpatient setting with enlarged lymph nodes as the most common symptom
- Most lymphoma cases definitively diagnosed through biopsy or flow cytometry
- Majority of cases will have imaging studies (e.g. CT, MRI, PET scans) to aid in diagnosis and determine extent of disease

## METHODS

### *Data Source and Study Population*

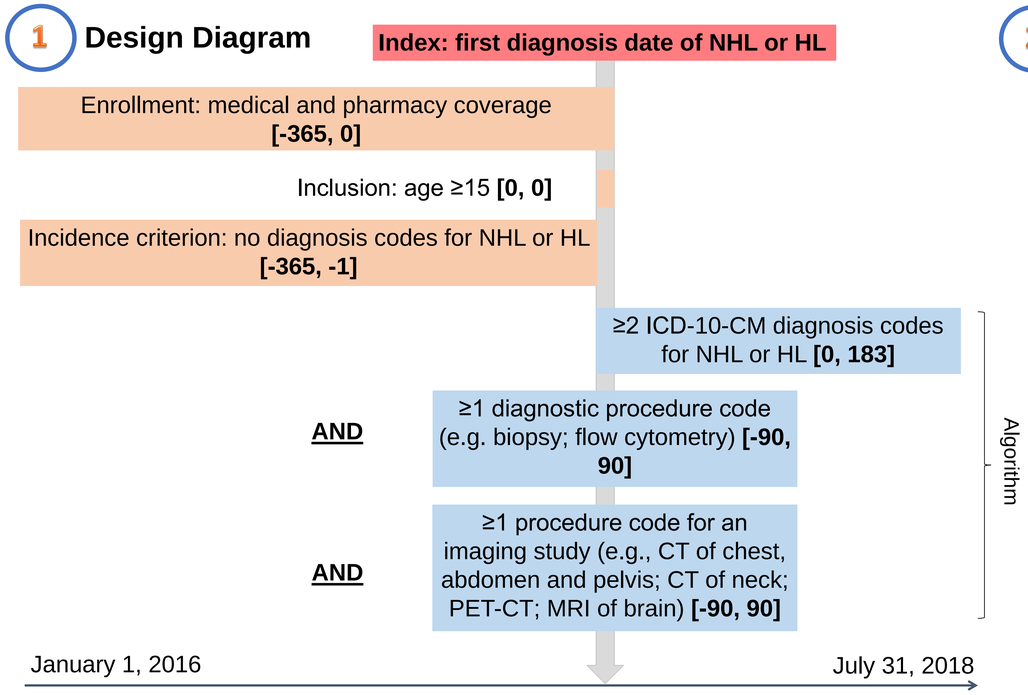
- Four Sentinel Data Partners (3 national insurers, 1 integrated healthcare system) contributed data
- Eligible participants were aged ≥15 years and enrolled for ≥12 months before diagnosis

### *Algorithm Development and Evaluation*

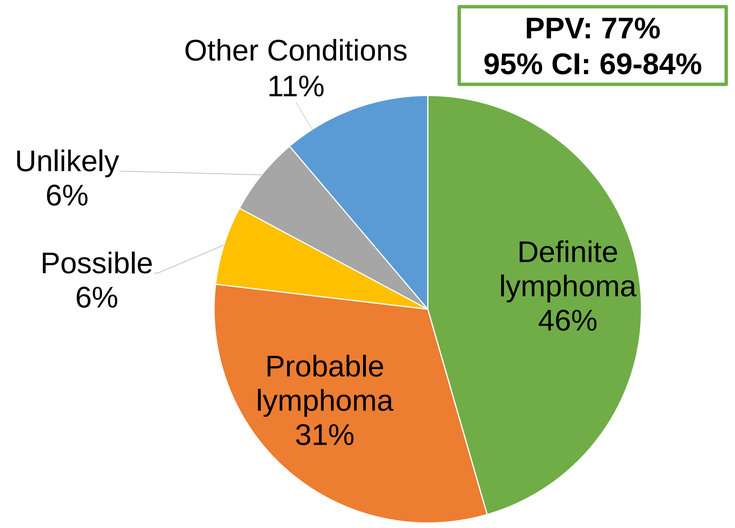
- Three component algorithm: 2 lymphoma-related diagnosis codes within 183 days, ≥1 diagnostic procedure and ≥1 relevant imaging code ± 90 days from first diagnosis (Figure 1)
- De-identified, patient-level claims data were extracted for 211 of the 8723 patients identified by the algorithm and reviewed by two oncologists to select encounters for chart retrieval
- 134 full charts from algorithm-positive cases were abstracted and adjudicated with data from -30 to +90 days from selected encounter date
- Subtype data (NHL vs HL) explored
- Definite and Probable cases considered true positives for calculation of the Positive Predictive Value (PPV)

## RESULTS

### 1 Design Diagram




### 2 Chart Adjudication Results (N = 134)




Cases judged as definite or probable lymphoma were considered true positives. Other conditions (N=15) included other cancers (80%) and non-malignant conditions (20%)

### 3 Demographic Characteristics by Chart Adjudication Status

| | | Final Adjudicated Case Status | | | | | Subtype | | |
|---|---|---|---|---|---|---|---|---|---|
| | Overall | Definite | Probable | Possible | No evidence | Other condition | NHL | HL | Non-case |
| **Total N (%)** | 134 | 61 (46%) | 42 (31%) | 8 (6%) | 8 (6%) | 15 (11%) | 92 (69%) | 11 (8%) | 31 (23%) |
| **Age, years** | | | | | | | | | |
| *Mean* | 62.2 | 62.4 | 63.7 | 70 | 52.6 | 58.1 | 65.7 | 39.7 | 59.8 |
| *Median* | 65.5 | 65 | 66.5 | 67.5 | 62 | 66 | 67 | 34 | 66 |
| *Range* | 17-94 | 21-85 | 17-83 | 49-94 | 18-79 | 18-90 | 17-85 | 17-72 | 18-94 |
| **Sex, N (%)** | | | | | | | | | |
| *Male* | 69 (52%) | 36 (59%) | 21 (50%) | 4 (50%) | 0 | 8 (53%) | 54 (59%) | 3 (27%) | 12 (39%) |
| *Female* | 64 (48%) | 25 (41%) | 20 (48%) | 4 (50%) | 8 (100%) | 7 (47%) | 37 (40%) | 8 (73%) | 19 (61%) |
| *Missing* | 1 (<1%) | 0 | 1 (2%) | 0 | 0 | 0 | 1 (1%) | 0 | 0 |

## DISCUSSION

### *Strengths*

- This project updates prior Sentinel work to develop an algorithm using ICD-10 codes
- Inclusion of data from both national insurers and an integrated delivery system enhanced generalizability of the results

### *Limitations*

- Chart reviews limited to one encounter
- Future iterations may add steps to further rule out non-lymphoma malignancies

### *Conclusions*

- An ICD-10-based algorithm including both diagnosis and procedure codes can identify lymphoma cases from health claims data with reasonable accuracy
- Subtype (NHL or HL) correctly determined for most cases

## ACKNOWLEDGEMENTS

- The authors would like to acknowledge the contributions of Yunping Zhou (Humana Healthcare Research, Inc. (HHR)), Jennifer Kuntz (Kaiser Permanente Northwest), Kevin Haynes, Lauren Parlett, and Shia Kent (HealthCore (Anthem)), and Michael Nguyen (FDA).
- This work was supported by the U.S. Food and Drug Administration contract: FDA HHSF223201400030I. MME was supported in part by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant KL2TR001454.
- The views expressed are those of the authors and are not to be construed as conveying either an official endorsement or criticism by U.S. Department of Health and Human Services, or U.S. Food and Drug Administration.
- Many thanks are due to Data Partners who provided data used in the analysis.