Published in final edited form as:

Ann Epidemiol. 2013 May ; 23(5): 291–293. doi:10.1016/j.annepidem.2013.02.005.

Accuracy of ICD-9CM Codes in Identifying Infections of Pneumonia and Herpes Simplex Virus in Administrative Data

Jennifer Drahos, PhD, MPH^{1,2}, Jeffrey J. Vanwormer, PhD³, Robert T. Greenlee, PhD, MPH³, Ola Landgren, MD⁴, and Jill Koshiol, PhD¹

¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, Maryland ²Cancer Prevention Fellowship Program, Division of Cancer Prevention, National Cancer Institute, NIH, DHHS, Bethesda, Maryland ³Marshfield Clinic Research Foundation, Marshfield, Wisconsin ⁴Multiple Myeloma Section, Metabolism Branch, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland

Introduction

Clinical epidemiology studies increasingly rely on electronic medical records data. The validity of ICD-9CM diagnosis codes is crucial as they are often used to identify conditions of interest. While many studies have used and validated ICD-9CM codes for chronic conditions, such as rheumatoid arthritis or cancer, few studies have utilized ICD-9CM codes to identify acute infectious conditions. Therefore we evaluated the utility of archived ICD-9CM codes to identify two representative infection-related conditions, pneumonia and herpes simplex virus (HSV), in a defined health system. In addition, we explored strategies to improve the standard ICD-9CM code-based selection.

Methods

Study Design

Using 2000 to 2010 data within a defined cohort of Marshfield Clinic patients and Security Health Plan [1], we identified potential cases with ICD-9CM diagnostic codes for pneumonia (480–486, N=25,064) or HSV (054.0–054.9, N=5,661). We selected an evaluation sample of 175 subjects with a pneumonia code and 179 subjects with an HSV infection code for validation via medical chart review and adjudication. For each subject, trained research coordinators reviewed the electronic medical records and completed disease-specific abstraction forms, collecting information about patient demographics, infection-related symptoms, laboratory test results, and prescribed medications. The study investigators adjudicated each subject as a case, non-case, or equivocal case (considered part of the non-case group for analytical purposes) by comparing abstracted data against pre-determined case definitions.

Corresponding author: Jennifer Drahos, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, 6120 Executive Blvd, EPS/Suite 550/Room 5034, Bethesda, MD 20852-7234, phone: 301-443-9444, fax: 301-402-0916, jennifer.drahos@nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Statistical analysis

Using SAS 9.2 [2] we calculated frequencies and percentages for categorical variables and compared demographic features between groups using chi-square or Fisher's exact tests, as necessary. A two-sided P value of <0.05 was considered statistically significant.

The positive predictive value (PPV) can be calculated directly as the probability of having the infectious condition given that there was positive confirmation by medical chart review. Specifically, PPV was calculated as the proportion of individuals correctly identified as a case based upon medical chart review among all individuals reviewed with an ICD-9CM code for pneumonia or HSV. We evaluated three additional ascertainment strategies to determine if 1) presence of two or more relevant ICD-9CM diagnosis codes, 2) ICD-9CM code plus the presence of a relevant prescription or 3) excluding those diagnosed during hospitalization could improve the accuracy of the ICD-9CM code to identify true cases without significant loss in case ascertainment. We could not evaluate sensitivity and specificity of case ascertainment using ICD-9CM codes since patients without diagnostic codes for pneumonia and HSV could not be reviewed, but instead calculated the proportion of true positives identified using the augmented strategy compared to true positives identified by the single ICD-9CM.

Results

ICD-9CM Validation

Pneumonia ICD-9CM codes confirmed medical record case-status in 88% of patients, however 10% were non-cases, and the evidence for 2% was equivocal. Similar results were observed for HSV, with 86% cases, 7% non-cases, and 7% equivocal. The presence of a single ICD-9CM code had a PPV of 88% for pneumonia and 86% for HSV. Demographic and clinical characteristics of the verified cases were compared to the non-cases and equivocal cases (Table I).

Verified pneumonia cases received multiple ICD-9CM codes for pneumonia more often than non-cases/equivocal cases (63.6% vs. 47.6%, $P=0.16$). In contrast, most HSV patients received only 1 ICD-9CM code (65.3%), regardless of case-status (66.8% vs. 56%, $P=0.29$). Physician prescription was part of the gold standard definition for pneumonia but not HSV; however, both pneumonia (96.8% vs. 3.2%, $p<0.01$) and HSV (92.9% vs. 7.1%, $p=0.01$) cases were more likely than non-cases/equivocal cases to receive an antibiotic prescription.

Additional Selection Strategies

Requiring an ICD-9CM code plus a documented prescription improved the PPV for pneumonia (88.0% to 96.8%), but decreased the PPV for HSV (86.0% to 76.6%). The percent of true cases identified with a single ICD-9CM code was high using this selection criteria for both pneumonia (99.4%) and HSV (92.9%). In contrast, requiring multiple ICD-9CM only modestly improved the PPV for pneumonia cases (88.0% to 90.7%) but dramatically decreased the number of true cases identified with a single ICD-9CM code (63.6%). This strategy did not improve the PPV of HSV (86.0% to 82.3%) and also resulted in lower capture of true HSV cases (33.1%). Lastly, excluding diagnoses from inpatients improved the PPV for both pneumonia (88.0% to 92.4%) and HSV (86.0% to 87.1%) and decreased percent of true pneumonia (87.0%) and HSV cases (95.1%). Several additional strategies tested decreased both PPV and capture of true cases (data not shown).

Discussion

Our study demonstrates that ICD-9CM codes for pneumonia and HSV was a valid marker of a true history of these conditions. False-positives (non-cases) accounted for < 10% of records evaluated for each condition. The PPVs for ICD-9CM codes may differ across health care settings (e.g., inpatient), an important consideration when selecting cases. Furthermore, requiring a documented prescription, in addition to the infection-specific ICD-9CM code, may improve correct identification of pneumonia cases. Although requiring multiple ICD-9CM codes in previous ICD-9CM validation studies increased specificity for other conditions [3–5], we found little improvement in the PPV for pneumonia and worse PPV for HSV by requiring multiple ICD-9CM codes. Strategies to improve the accuracy of ICD-9CM codes are likely infection-specific.

To our knowledge this study is the first to verify the accuracy of ICD-9CM codes in a multispecialty healthcare system to identify pneumonia or adult HSV. Previous pneumonia ICD-9CM validation studies focused on veterans, hospitalized patients, or the elderly [6–10], while the accuracy of ICD-9CM codes for HSV has only been assessed for the identification of neonates [11, 12]. A strength of our study is that it encompassed patient visits at multiple facilities within the healthcare system including physician, emergency, and hospital visits. However, we could not evaluate patients without ICD-9CM codes for these conditions and therefore could not determine the sensitivity of the ICD-9CM codes. Although the results may not be generalizable in health care settings with different documentation or ICD-9CM coding practices, our results were similar to previous reports from inpatient settings.

Validation studies of ICD-9CM codes may help identify potential administrative documentation gaps, and incorporating other medical record data, especially prescription data, may improve ICD-9CM code performance. That said, our results suggest that ICD-9CM codes can be used to successfully identify infection-related conditions in epidemiologic studies.

Acknowledgments

The authors would like to thank the study team at Marshfield Clinic Research Foundation for their work on this project: Nick Berger, Marilyn Bruger, Deanna Cole, Autumn Deedon, Deborah Hilgemann, Paul Hitz, Deb Johnson, Tara Johnson, Deb Kempf, Diane Kohnhorst, Cyndy Meyer, Aaron Miller, Suellen Murray, DeAnn Polacek, Katie Pralle, Theresa Pritzl, Ashley Quinnell, Kristina Reisner, Sandy Strey, Rachelle Tuyls, and Daphne York.

References

1. DeStefano F, et al. Epidemiologic research in an integrated regional medical care system: the Marshfield Epidemiologic Study Area. *J Clin Epidemiol.* 1996; 49(6):643–52. [PubMed: 8656225]
2. SAS Institute Inc., SAS 9.2. Cary, NC: SAS Institute Inc;
3. Molodecky NA, et al. Validity of administrative data for the diagnosis of primary sclerosing cholangitis: a population-based study. *Liver Int.* 2011; 31(5):712–20. [PubMed: 21457444]
4. Harrold LR, et al. Validity of gout diagnoses in administrative data. *Arthritis Rheum.* 2007; 57(1): 103–8. [PubMed: 17266097]
5. Stein BD, et al. The Validity of ICD-9-CM Diagnosis Codes for Identifying Patients Hospitalized for COPD Exacerbations. *Chest.* 2011
6. Aronsky D, et al. Accuracy of administrative data for identifying patients with pneumonia. *Am J Med Qual.* 2005; 20(6):319–28. [PubMed: 16280395]

7. Guevara RE, et al. Accuracy of ICD-9-CM codes in detecting community-acquired pneumococcal pneumonia for incidence and vaccine efficacy studies. *Am J Epidemiol.* 1999; 149(3):282–9. [PubMed: 9927225]
8. Skull SA, et al. ICD-10 codes are a valid tool for identification of pneumonia in hospitalized patients aged > or = 65 years. *Epidemiol Infect.* 2008; 136(2):232–40. [PubMed: 17445319]
9. van de Garde EM, et al. International classification of diseases codes showed modest sensitivity for detecting community-acquired pneumonia. *J Clin Epidemiol.* 2007; 60(8):834–8. [PubMed: 17606180]
10. Yu O, et al. Classification algorithms to improve the accuracy of identifying patients hospitalized with community-acquired pneumonia using administrative data. *Epidemiol Infect.* 2011; 139(9): 1296–306. [PubMed: 21087538]
11. Flagg EW, Weinstock H. Incidence of neonatal herpes simplex virus infections in the United States, 2006. *Pediatrics.* 2011; 127(1):e1–8. [PubMed: 21149432]
12. Xu F, et al. Incidence of neonatal herpes simplex virus infections in two managed care organizations: implications for surveillance. *Sex Transm Dis.* 2008; 35(6):592–8. [PubMed: 18418296]

Table 1
Demographic and Clinical Characteristics of Verified Cases and Non-Cases & Equivocal Cases

	Pneumonia (n=175)			Herpes (n=179)		
	Case	Non-Case & Equivocal	P-value	Case	Non-Case & Equivocal	P-value
Total	154 (88.0%)	21 (12.0%)		154 (86.0%)	25 (14.0%)	
Sex						
Females	70 (45.4%)	13 (61.9%)	0.27	109 (70.7%)	20 (80.0%)	0.40
Males	84 (54.5%)	8 (38.0%)	0.37	45 (29.2%)	5 (20.0%)	0.66
Age Group						
child (<=19)	73 (47.4%)	3 (14.2%)	0.26	51 (33.1%)	5 (20.0%)	0.55
adult (20–59)	36 (23.3%)	9 (42.8%)	0.24	74 (48.0%)	16 (64.0%)	0.25
senior (60+)	45 (29.2%)	9 (42.8%)	0.42	29 (18.8%)	4 (16.0%)	0.92
RUCA Score						
Urban	30 (19.4%)	3 (14.2%)	0.83	27 (17.5%)	5 (20.0%)	0.89
Rural	124 (80.5%)	18 (85.7%)	0.60	127 (82.4%)	20 (80.0%)	0.80
Physician Type						
Family/Internal Medicine	63 (40.9%)	7 (33.3%)	0.70	58 (37.6%)	11 (44.0%)	0.69
Pediatrician	38 (24.6%)	0 (0.0%)		20 (12.9%)	1 (4.0%)	0.47
Specialist	9 (5.84%)	8 (38.0%)	0.10	31 (20.1%)	8 (32.0%)	0.57
Emergency/Urgent Care	28 (18.1%)	2 (9.5%)	0.76	29 (18.8%)	0 (0.0%)	0.02
Other/Unknown	16 (10.3%)	4 (19.0%)	0.63	16 (10.3%)	5 (20.0%)	0.82
Visit Setting						
Outpatient	118 (76.6%)	7 (33.3%)	0.01	137 (88.9%)	17 (68.0%)	0.02
Emergency	16 (10.3%)	2 (9.5%)	0.97	6 (3.9%)	2 (8.0%)	0.82
Inpatient	20 (12.9%)	10 (47.6%)	0.04	6 (3.9%)	3 (12.0%)	0.64
Lab/Imaging Only	0 (0%)	2 (9.6%)		5 (3.2%)	3 (12.0%)	0.63
Total ICD-9 Codes						
1	56 (36.3%)	11 (52.3%)	0.32	103 (66.8%)	14 (56.0%)	0.43
2+	98 (63.6%)	10 (47.6%)	0.32	51 (33.1%)	11 (44.0%)	0.49
Prescription						
Yes	153 (99.3%)	5 (23.8%)	<0.01	118 (76.6%)	9 (36.0%)	0.01

	Pneumonia (n=175)			Herpes (n=179)		
	Case	Non-Case & Equivocal	P-value	Case	Non-Case & Equivocal	P-value
Infection-Specific Test ²						
No	1 (0.64%)	16 (76.1%)	.	36 (23.3%)	16 (64.0%)	<0.01
Yes	129 (83.7%)	15 (71.4%)	0.24	33 (21.4%)	5 (20.0%)	0.94
No	25 (16.2%)	6 (28.5%)	0.49	121 (78.5%)	20 (80.0%)	0.88

¹ Percentages provided across column for all categories

² Imaging with X-ray or CT scan for pneumonia infection or PCR, ELISA, or culture for Herpes simplex virus infection