

ORIGINAL ARTICLE

Efficiency of *International Classification of Diseases, Ninth Revision*, Billing Code Searches to Identify Emergency Department Visits for Blood or Body Fluid Exposures through a Statewide Multicenter Database

Lisa M. Rosen, ScM;^{1,2} Tao Liu, PhD;¹ Roland C. Merchant, MD, MPH, ScD^{1,2}

BACKGROUND. Blood and body fluid exposures are frequently evaluated in emergency departments (EDs). However, efficient and effective methods for estimating their incidence are not yet established.

OBJECTIVE. Evaluate the efficiency and accuracy of estimating statewide ED visits for blood or body fluid exposures using *International Classification of Diseases, Ninth Revision (ICD-9)*, code searches.

DESIGN. Secondary analysis of a database of ED visits for blood or body fluid exposure.

SETTING. EDs of 11 civilian hospitals throughout Rhode Island from January 1, 1995, through June 30, 2001.

PATIENTS. Patients presenting to the ED for possible blood or body fluid exposure were included, as determined by prespecified ICD-9 codes.

METHODS. Positive predictive values (PPVs) were estimated to determine the ability of 10 ICD-9 codes to distinguish ED visits for blood or body fluid exposure from ED visits that were not for blood or body fluid exposure. Recursive partitioning was used to identify an optimal subset of ICD-9 codes for this purpose. Random-effects logistic regression modeling was used to examine variations in ICD-9 coding practices and styles across hospitals. Cluster analysis was used to assess whether the choice of ICD-9 codes was similar across hospitals.

RESULTS. The PPV for the original 10 ICD-9 codes was 74.4% (95% confidence interval [CI], 73.2%–75.7%), whereas the recursive partitioning analysis identified a subset of 5 ICD-9 codes with a PPV of 89.9% (95% CI, 88.9%–90.8%) and a misclassification rate of 10.1%. The ability, efficiency, and use of the ICD-9 codes to distinguish types of ED visits varied across hospitals.

CONCLUSIONS. Although an accurate subset of ICD-9 codes could be identified, variations across hospitals related to hospital coding style, efficiency, and accuracy greatly affected estimates of the number of ED visits for blood or body fluid exposure.

Infect Control Hosp Epidemiol 2012;33(6):581-588

The *International Classification of Diseases, Ninth Revision (ICD-9)*, codes are used frequently in research studies for database searches. ICD-9 is a standardized coding system used by hospitals to classify procedures and diagnoses. Some authors have reported high positive predictive values (PPVs) when using ICD codes to identify patients and patient visits associated with particular diagnoses in databases.¹⁻⁴ However, other authors have criticized this search method, citing evidence that ICD coding errors in databases can lead to biased estimates of the frequency of visits for or the prevalence of some medical conditions.⁵⁻⁷ ICD coding errors result from coding entry mistakes, problems with coders interpreting

medical records, errors in diagnosis and medical record documentation, miscommunication between patients and healthcare providers, inaccurate test results, and variability in clinical terminology among healthcare providers.^{8,9} Using ICD codes in combination with another source of information can increase the accuracy with which particular diagnoses are identified.¹⁰ Medical record reviews are commonly employed to verify the results of ICD code searches of databases; however, record reviews can be time consuming and costly. It would be helpful to know which types of medical visits or conditions are amenable to an ICD code search without medical record verification before initiating studies involving large

Affiliations: 1. Department of Community Health, Alpert Medical School of Brown University, Providence, Rhode Island; 2. Department of Emergency Medicine, Alpert Medical School of Brown University, Providence, Rhode Island.

Received August 2, 2011; accepted January 4, 2012; electronically published April 12, 2012.

© 2012 by The Society for Healthcare Epidemiology of America. All rights reserved. 0899-823X/2012/3306-0007\$15.00. DOI: 10.1086/665722

databases, performing multicenter investigations, and conducting studies on the frequency of visits for or the prevalence of specific conditions.

We previously conducted a large statewide multicenter study on the incidence of emergency department (ED) visits for blood or body fluid exposures (ie, occupational and non-occupational exposures to infectious diseases, particularly human immunodeficiency virus infection and hepatitis).¹¹⁻¹⁸ In the earlier study, we employed an ICD-9 search of billing records followed by a direct medical record review to verify ED visits as being for blood or body fluid exposures. Our goal in this study was to determine whether the incidence of ED visits for blood or body fluid exposures might be ascertained more efficiently, such as through an ICD-9 code search without a direct medical record review. We wanted to assess whether a subset of ICD-9 codes could be used to increase the efficiency of these searches while maintaining the accuracy of consequent estimates of the incidence of ED visits for these exposures. We also aimed to examine whether variations in the use of ICD-9 codes across hospitals, which might reflect differences in coding practices and styles, might influence estimates of the incidence of ED visits for blood or body fluid exposure. If these variations could be explained or anticipated, then estimates of ED visits for blood or body fluid exposure through an ICD-9 code search across multiple institutions could be adjusted using mathematical models to improve the accuracy of the estimates. If a more efficient method of determining the incidence of ED visits for blood or body fluid exposure were possible, then this method could be employed at other locations to assist in planning for services and resources for patients who present for care for these exposures.

METHODS

Study Design

This study was a secondary analysis of a database previously used to estimate the incidence of ED visits for blood or body fluid exposure.¹² The study was approved by the hospital's institutional review board.

Study Database

The methods for the original study have been previously described and are briefly summarized here.¹² Data were collected from 11 civilian hospitals throughout the state for patients presenting to EDs for possible blood or body fluid exposure from January 1, 1995, through June 30, 2001. Blood or body fluid exposure was defined as percutaneous injuries, blood or body fluid splashes, and sexual contact (consensual sex or sexual assault). Participating hospitals searched their hospital billing databases and their ED practitioner billing databases (when available) using 10 ICD-9 codes (Table 1). An earlier pilot study determined these codes to be those most likely to identify ED visits for blood or body fluid exposure.¹¹ The

TABLE 1. Description of *International Classification of Diseases, Ninth Revision (ICD-9)*, Codes

ICD-9 code	Diagnosis
995.53	Child sexual abuse
995.83	Adult sexual abuse
V01.7	Other viral diseases
V01.8	Other communicable diseases
V07.8	Other specified prophylactic measure
V07.9	Unspecified prophylactic measure
V15.41	History of physical abuse
V15.85	Exposure to potentially hazardous body fluids
V71.5	Observation following alleged rape or seduction
E920.5	Hypodermic needle

ICD-9 codes could be for primary, secondary, or tertiary diagnoses.

Trained reviewers extracted data from the medical records onto a standardized form, and ED visits were categorized as being for blood or body fluid exposure or not for blood or body fluid exposure. Visits for other medical conditions that could have constituted correct use of the ICD-9 codes employed in the search (eg, use of V01.8 for varicella exposure) and visits that were coded incorrectly (ie, for which a different code should have been used) were categorized as not being for blood or body fluid exposure. Medical records that could not be located or records without extractable data (ie, blank or incomplete records) were considered to be missing. One of the 11 hospital billing databases had no records from earlier than 1998 to review, which prevented data from these charts from being extracted. The ICD-9 code search revealed 5,153 ED visits, of which 3,639 (70.6%) were for blood or body fluid exposure and 1,250 (24.3%) were not for blood or body fluid exposure; there were 264 (5.1%) missing medical records that were excluded from the analysis.

Data Analysis

Efficiency of ICD-9 code search in identifying ED visits for blood or body fluid exposure. The PPVs for each of the 10 ICD-9 codes in identifying an ED visit for blood or body fluid exposure were estimated with corresponding 95% confidence intervals (CIs). The proportion of ED visits for blood or body fluid exposure that would not have been identified if each ICD-9 code were excluded from the database search was calculated along with corresponding 95% CIs. The data were divided into a training set ($n = 3,260$ visits) and a test set ($n = 1,629$ visits). Recursive partitioning,¹⁹ which successively divides the data into homogenous groups, was used on the training set to determine which subset of the 10 ICD-9 codes would most efficiently identify ED visits for blood or body fluid exposure. By sequentially examining subsets of ICD-9 codes, we looked for the best trade-off that identified the most ED visits for blood or body fluid exposure while achieving the best PPV and lowest misclassification rate. The

TABLE 2. Frequency with which *International Classification of Diseases, Ninth Revision (ICD-9)*, Codes Distinguished Emergency Department (ED) Visits for Blood or Body Fluid Exposure from ED Visits not for Blood or Body Fluid Exposure in the Database

ICD-9 code or code combination	No. of visits in database	ED visit for blood or body fluid exposure as identified by ICD-9 code and code combinations, no. (%) of ED visits	
		No	Yes
995.53	116	18 (15.5)	98 (84.5)
995.83	37	2 (5.4)	35 (94.6)
E920.5	774	78 (10.1)	696 (89.9)
V01.7	211	170 (80.6)	41 (19.4)
V01.8	386	350 (90.7)	36 (9.3)
V07.8	299	285 (95.3)	14 (4.7)
V07.9	10	9 (90.0)	1 (10.0)
V15.41	20	17 (85.0)	3 (15.0)
V15.85	908	107 (11.8)	801 (88.2)
V71.5	1,597	173 (10.8)	1,424 (89.2)
V07.8, V01.7	9	9 (100.0)	0 (0.0)
V15.85, V01.7	11	1 (9.1)	10 (90.9)
V15.85, V01.8	3	1 (33.3)	2 (66.7)
E920.5, V15.85	322	9 (2.8)	313 (97.2)
995.53, V71.5	131	5 (3.8)	126 (96.2)
V07.8, V15.85	4	1 (25.0)	3 (75.0)
995.83, V71.5	26	2 (7.7)	24 (92.3)
V01.7, V01.8	8	6 (75.0)	2 (25.0)
V71.5, V01.8	1	0 (0.0)	1 (100.0)
V07.8, V07.9	1	1 (100.0)	0 (0.0)
V07.8, V01.8	4	4 (100.0)	0 (0.0)
E920.5, V01.8	3	1 (33.3)	2 (66.7)
V71.5, V15.85	1	0 (0.0)	1 (100.0)
E920.5, V01.7	3	0 (0.0)	3 (100.0)
995.53, 995.83, V71.5	1	0 (0.0)	1 (100.0)
V01.7, V01.8, V07.8	1	1 (100.0)	0 (0.0)
V01.7, E920.5, V15.85	1	0 (0.0)	1 (100.0)
V71.5, V15.85, 995.53	1	0 (0.0)	1 (100.0)
Total	4,889	1,250	3,639

test set was used to assess the predictive ability of the ICD-9 codes chosen from the recursive partitioning analysis.

Variations in efficiency of the yield of ICD-9 code searches for ED visits for blood or body fluid exposure across hospitals. We examined the efficiency of the yield of ICD-9 code searches across hospitals. A search would be considered less efficient if ED visits that were not for blood or body fluid exposure were identified more frequently at a given hospital than at other hospitals. Pearson χ^2 test was used to compare the frequency distribution of ED visits that were not for blood or body fluid exposure across institutions.

Random-effects logistic regression modeling, which takes into account the hierarchical structure of the data, was conducted to identify institution-level factors that might explain differences in the efficiency of the yield of ED visits for blood or body fluid exposure identified through the ICD-9 code search. The regression was adjusted for each hospital (the random effect term) and for the following institution-level

factors (fixed effects): hospital type (academic or community), hospital location (urban or nonurban),²⁰ type of billing database (hospital or hospital and ED practitioner), the number of patient beds in the hospital, and year. The number of patient beds in the hospital was used as a proxy for hospital size and patient volume, and year was included as a covariate to adjust for any temporal trends that may have affected the number of patients presenting to the ED for blood or body fluid exposure. We also examined whether the institution-level factors accounted for variations in identification of ED visits that were not for blood or body fluid exposure across hospitals. In other words, we tested whether the variance of the random effect was equal to 0 using a likelihood ratio test.²¹ The above analysis was repeated using the subset of ICD-9 codes identified in the recursive partitioning analysis.

Variations in the use of ICD-9 codes for ED visits for blood or body fluid exposure across hospitals. We performed a cluster analysis, using complete linkage, to assess whether the use

TABLE 3. Positive Predictive Value (PPV) of Individual *International Classification of Diseases, Ninth Revision* (ICD-9), Codes for Blood or Body Fluid Exposure Emergency Department (ED) Visits and Visits Missed because of Exclusion of Codes from Search

ICD-9 code	PPV for ED visits for blood or body fluid exposure (95% CI)	Proportion of ED visits for blood or body fluid exposure missed if code was excluded from database search, % (95% CI) ^a
Other specified prophylactic measure (V07.8)	5.4 (3.4–8.4)	0.4 (0.2–0.6)
Unspecified prophylactic measure (V07.9)	9.1 (1.6–37.7)	0.0 (0.0–0.2)
Other communicable diseases (V01.8)	10.6 (8.0–14.0)	1.0 (0.7–1.4)
History of physical abuse (V15.41)	15.0 (5.2–36.0)	0.1 (0.0–0.2)
Other viral diseases (V01.7)	23.4 (18.5–29.1)	1.1 (0.8–1.5)
Observation following alleged rape or seduction (V71.5)	89.8 (88.3–91.1)	39.1 (37.6–40.7)
Exposure to potentially hazardous body fluids (V15.85)	90.5 (88.7–92.0)	22.0 (20.7–23.4)
Child sexual abuse (995.53)	90.8 (86.5–93.8)	2.7 (2.2–3.3)
Hypodermic needle (E920.5)	92.0 (90.3–93.5)	19.1 (17.9–20.4)
Adult sexual abuse (995.83)	93.8 (85.0–97.5)	1.0 (0.7–1.3)

NOTE. CI, confidence interval.

^a Proportions do not total to 100% because some visits could be captured by another ICD-9 code.

of ICD-9 codes was similar across certain institutions. Cluster analysis is a descriptive and exploratory analysis that groups similar objects together into “clusters.” Euclidean distance was used to quantify the dissimilarity between institutions.²² We standardized the measure of ED visits across hospitals by dividing the number of visits associated with each ICD-9 code or combination of codes by the total number of observed ED visits for that hospital. Because visits for blood or body fluid exposure varied by ED (eg, some hospitals were more likely to have ED visits for sexual assault), we performed a stratified cluster analysis using the 5 most common types of blood or body fluid exposure: adult female sexual assault, pediatric sexual assault, healthcare worker blood or body fluid exposure, adult nonhealthcare worker blood or body fluid exposure, and pediatric nonhealthcare worker blood or body fluid exposure.

All statistical analyses were conducted in R, version 2.7.1, and Stata 11 (Stata). Differences were considered to be significant at the $\alpha = .05$ level.

RESULTS

Efficiency of ICD-9 Code Search in Identifying ED Visits for Blood or Body Fluid Exposure

Table 2 depicts how frequently each ICD-9 code or combination of codes appeared in the final database from the search for ED visits for blood or body fluid exposure across the 11 hospitals. Also portrayed are the proportions of ED visits by ICD-9 code that were verified through the medical record review as ED visits for a blood or body fluid exposure. As shown, the yield of each ICD-9 code or combination of ICD-9 codes in correctly identifying these as ED visits for blood or body fluid exposure varied greatly. When the PPVs for each of the 10 ICD-9 codes were evaluated individually, 5 codes had PPVs of approximately 90% or greater, whereas the remainder had PPVs of less than 25% (Table 3). The

overall PPV for distinguishing ED visits for blood or body fluid exposure from ED visits that were not for blood or body fluid exposure for all 10 ICD-9 codes (individually and in combination) was 74.4% (95% CI, 73.2%–75.7%).

For each of the 10 ICD-9 codes, Table 3 also provides the proportions of ED visits for blood or body fluid exposure in the final database that would not have been identified if each code had not been used in the search for these visits. For the 5 ICD-9 codes with PPVs less than 25%, the proportion of ED visits for blood or body fluid exposure that would have been missed if these codes were excluded was low (0%–1.1%). However, even though the PPVs for codes 995.53 (child sexual abuse) and 995.83 (adult sexual abuse) were high, few ED visits for blood or body fluid exposure would have been missed if these codes had been excluded from the search. These 2 ICD-9 codes were commonly used in combination with V71.5 (observation following alleged rape or seduction).

Figure 1 depicts the results of the recursive partitioning analysis that aimed to find an optimal subset of ICD-9 codes that could be used in a more efficient search for ED visits for blood or body fluid exposure. Node 1 of the resultant tree shows the overall misclassification rate (25.6%) for identifying ED visits for blood or body fluid exposure from the training set if all 10 codes were used. Nodes 2, 4, and 5 represent possible subsets of the 5 codes V01.7, V01.8, V07.8, V07.9, V15.41, and any of their combinations. The subset of codes at nodes 2, 4, and 5 have high misclassification rates (>80%) and include only 67 (2.8%) or fewer ED visits for blood or body fluid exposure. As noted in Table 3, these 5 codes also had low PPVs (<25%) and missed few ED visits (<1.1%) if they were excluded from the database. The combination of ICD-9 codes at node 7 had the highest PPV (96.4%) and the lowest misclassification rate (3.6%); however, using the subset of codes at this node would result in missing 2,132 ED visits for blood or body fluid exposure (87.9%).

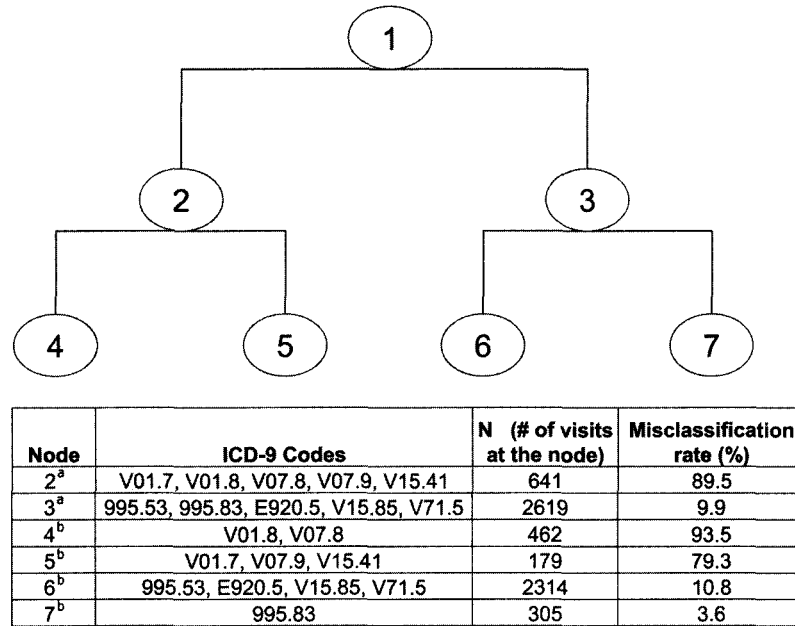


FIGURE 1. Regression tree analysis of *International Classification of Diseases, Ninth Revision (ICD-9)*, codes for blood or body fluid exposure emergency department visits. ^aAlso includes all combinations of these ICD-9 codes. ^bAlso includes a subset of combination codes from its parent node.

The subset of 5 codes at node 3 (V71.5, V15.85, 995.53, E920.5, 995.83, and any of their combinations) yielded a high PPV (90.1%; 95% CI, 88.9%–91.2%) and a low misclassification rate (9.9%), and few ED visits for blood or body fluid exposure were missed (67 visits), compared with all other nodes. These 5 codes were used to predict blood or body fluid exposure visits in the test set, which yielded consistent results (PPV: 89.6%; misclassification rate: 10.4%).

Variations in Efficiency of the Yield of ICD-9 Code Searches for ED Visits for Blood or Body Fluid Exposure across Hospitals

There was considerable variation across hospitals in the efficiency of the 10 ICD-9 codes in distinguishing ED visits for blood or body fluid exposure from other ED visits (Table 4). Hospitals F, I, J, and K had very different proportions of ED visits that were not for blood or body fluid exposure ($P < .001$, by Pearson χ^2 test), whereas only approximately 25% of ED visits identified in the ICD-9 code search for the remaining 7 hospitals were ultimately not for blood or body fluid exposure (per the medical record review). Table 5 summarizes the results of the random-effects logistic regression models that examined institution-level factors related to the efficiency of the yield of ICD-9 code searches in identifying ED visits for blood or body fluid exposure across hospitals. As shown in the results for all 10 ICD-9 codes (model 1), hospitals that demonstrated greater efficiency in the ICD-9 code search (ie, fewer ED visits identified that were not for

blood or body fluid exposure) were those located in an urban area and with dual (hospital and ED practitioner) billing databases. However, as indicated by the results of the likelihood ratio test ($P < .0001$), there remained considerable variation across hospitals (the random effect in the model) even when accounting for institution-level factors. In addition, when the subset of 5 ICD-9 codes that were suggested as being more efficient by the recursive partitioning analysis (node 3 of Figure 1) were examined (model 2), variation by hospital still existed ($P < .0001$).

TABLE 4. Emergency Department (ED) Visits Not for Blood or Body Fluid Exposure Identified through the *International Classification of Diseases, Ninth Revision*, Hospital Billing Database Search, by Hospital

Hospital	No. of ED visits	ED visits not for blood or body fluid exposure, %
Hospital A	341	23.5
Hospital B	133	24.1
Hospital C	1,224	24.0
Hospital D	283	29.0
Hospital E	182	27.5
Hospital F	411	42.1
Hospital G	865	26.4
Hospital H	146	28.1
Hospital I	470	11.7
Hospital J	549	19.9
Hospital K	285	37.2
Overall	4,889	25.6

TABLE 5. Random-Effects Logistic Regression Analysis of Institution-Level Factors Associated with Efficiency of *International Classification of Diseases, Ninth Revision (ICD-9)*, Code Search for Emergency Department (ED) Visits for Blood or Body Fluid Exposure

Institution-level factor	OR (95% CI)	
	Model 1 ^a	Model 2 ^b
Hospital type		
Community	Reference	Reference
Academic	1.53 (0.89, 2.61)	1.82 (0.97, 3.41)
Hospital location		
Nonurban area	Reference	Reference
Urban area	0.47 (0.27, 0.81)	0.73 (0.38, 1.39)
Presence of dual databases		
Not present	Reference	Reference
Present	0.59 (0.42, 0.83)	0.65 (0.39, 1.06)
No. of hospital beds ^c	1.02 (0.90, 1.17)	1.06 (0.91, 1.24)
Year		
1995	Reference	Reference
1996	0.96 (0.69, 1.33)	0.79 (0.45, 1.40)
1997	1.70 (1.22, 2.36)	0.87 (0.49, 1.54)
1998	1.31 (0.94, 1.81)	1.16 (0.67, 2.02)
1999	1.00 (0.72, 1.39)	1.65 (0.97, 2.80)
2000	0.81 (0.57, 1.15)	1.28 (0.74, 2.22)
2001	1.82 (1.29, 2.57)	1.41 (0.79, 2.52)

NOTE. CI, confidence interval; OR, odds ratio.

^a Model using all 10 ICD-9 codes.

^b Model using ICD-9 codes V71.5, V15.85, 995.53, E920.5, 995.83, and combinations thereof.

^c Results reported in units of 100 beds.

Variations in ICD-9 Code Use for ED Visits for Blood or Body Fluid Exposure across Hospitals

Figure 2 shows the results of the cluster analysis that assessed the choice of ICD-9 codes for ED visits across the 11 hospitals. The cluster dendrogram appears to have 4 clusters; however, there does not appear to be any relationship among the hospitals grouped according to institutional-level factors, such as hospital location, size, type of hospital, or use of administrative billing databases. In other words, the coding practices of each hospital appear to be distinct from each other and not based upon any clear pattern that can account for variations in coding practices. The results of the stratified cluster analyses by type of blood or body fluid exposure (adult female sexual assault, child sexual assault, healthcare worker blood or body fluid exposure, adult nonhealthcare worker blood or body fluid exposure, and child nonhealthcare worker blood or body fluid exposure) also do not indicate any observable pattern by institutional-level factors (Figure 2).

DISCUSSION

The overall goal of this study was to examine the efficiency of using the ICD-9 codes employed in an earlier study that estimated the incidence of ED visits for blood or body fluid exposure and to determine whether the results of this study

could indicate a better or easier method of conducting future multicenter studies on this topic. We found that employing the original set of 10 ICD-9 codes resulted in a 25.6% misclassification rate and that employing a subset of these codes resulted in missing a moderate number of cases but reduced the misclassification rate to 10.1%. Although the misclassification rate was significantly reduced, it was not negligible. As a result, forgoing a medical record review after an ICD-9 code search might overestimate the incidence of ED visits for these exposures. Overestimation can result from the use of general ICD-9 codes for multiple medical conditions (eg, the use of V01.8, "other communicable diseases") or the incorrect coding of visits.

We had hoped that perhaps we might be able to use institution-level factors related to the hospitals involved (eg, hospital type and number of patient beds) as covariates in a model that could assist in determining how to adjust any future estimates of the incidence of blood or body fluid exposure from an ICD-9 code-based incidence estimation. Level 1 covariates (ie, factors related to an individual's ED visit) were not considered, because these factors would require individual-level information obtained from additional data extraction. Variations across hospitals in regards to coding styles and practices appear to be quite marked, and we could not identify institution-level factors that might fully account for these variations. Hospitals that had both a hospital and an ED practitioner's billing database were less likely to have visits that were unrelated to blood or body fluid exposure. Combining these 2 data sets increased the number of visits for blood or body fluid exposure, which demonstrated that there are significant differences in how hospitals and ED practitioner coding groups review identical ED visits. Differences in coding between hospitals and ED practitioners should be further explored in future studies. Earlier studies have found that interrater agreement for ICD-9 coding is low for various diagnoses and conditions, such as psychiatric disorders and shoulder instability.^{23,24} Low interrater agreement may explain why the coding choice across hospitals appeared to be random. A limitation of our study was that the 11 institutions did not evaluate the same patients, and coders did not review the same medical charts. Therefore, we could not compute or analyze interrater agreement directly. However, we did analyze the data based on the population that presented to the ED. Because ICD-9 codes are selected to categorize patient symptoms and diagnosis, patients who present to the ED for the same type of blood or body fluid exposure should be comparable. Therefore, our stratified cluster analyses by patient population acted as a proxy for interrater agreement.

On the basis of our findings, we recommend that possible differences in coding style between specialty hospitals and other hospitals should be assessed before data collection. We identified no observable patterns across hospitals with regard to coding choice, which indicated that either coding style is random and dependent upon the specific hospital or there

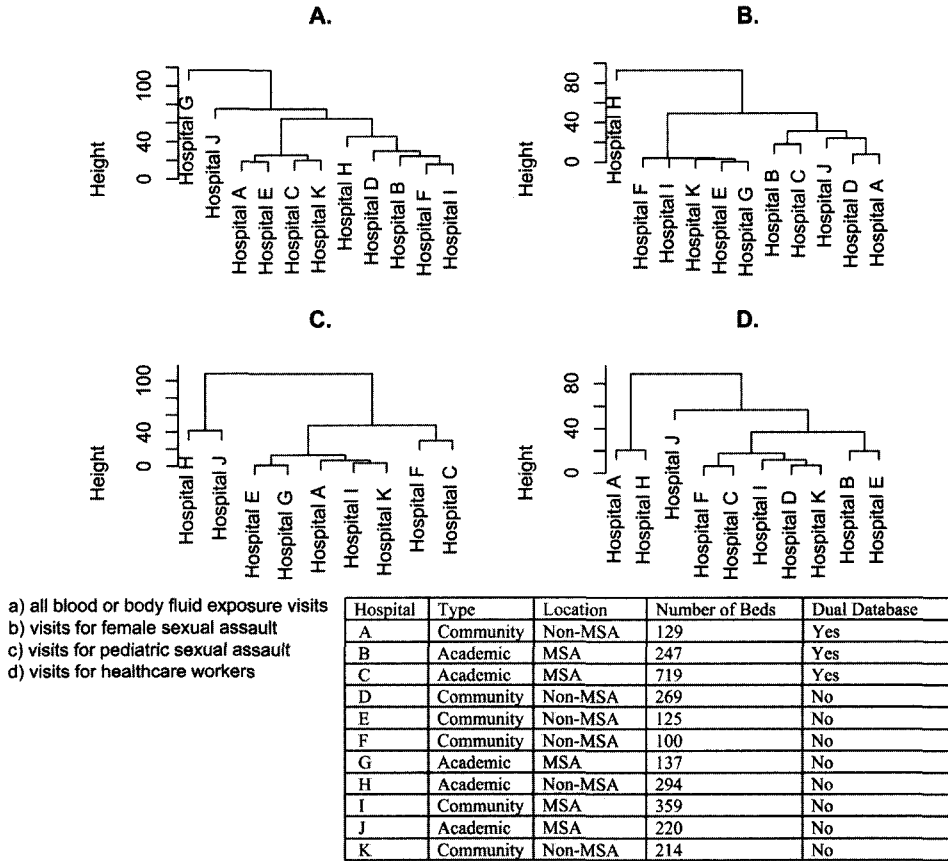


FIGURE 2. Cluster dendrogram of *International Classification of Diseases, Ninth Revision*, coding use across hospitals for emergency department visits for blood or body fluid exposure. MSA, metropolitan statistical area.

are individual-level variables and other influential variables that we cannot measure or observe. When selecting ICD-9 codes to identify blood or body fluid exposure visits at a different set of hospitals, researchers need to be aware that the codes used in a particular institution may not be the same as the codes used in a different institution, even if patients are presenting for the same reason. Therefore, it is in the researcher's best interest to consult with the coders at a hospital before searching the billing database.

Our research study has some limitations. First, we were unable to account for ED visits for blood or body fluid exposure that were coded with an ICD-9 code other than our 10 preselected codes. This limitation implies that the incidence of blood or body fluid exposure could be underestimated if there are additional visits that employed ICD-9 codes not included in our database. Second, the initial study design did not allow us to calculate other measures to evaluate the performance of ICD-9 codes, such as sensitivity and specificity. Third, there could have been other institutional-level variables or patient information that we could not measure with our data set, and therefore we are unable to account for these variables in our regression models. Fourth, we worked under the assumption that our missing data was missing com-

pletely at random. However, we have no reason to presume that missing charts were related to our outcome of interest or the selection of ICD-9 codes. Fifth, the data set is based only on the use of ICD-9 codes related to blood or body fluid exposure at the 11 hospitals included in this study and is not necessarily applicable to other data sets, ICD-9 codes, or institutions. However, the methods and procedures presented can be used in future studies by other researchers.

CONCLUSION

This study examined the use and efficiency of ICD-9 codes in identifying ED visits for blood or body fluid exposure as well as variation in the identification of these visits when data are obtained from multiple institutions. Although we found a more accurate subset of codes for visits at our 11 hospitals that had a high PPV and missed few relevant visits, we also observed variations in the proportion of ED visits that were not for blood or body fluid exposure that could not be explained by institutional factors and that seemed to be largely related to per-hospital coding style. However, there were no observable patterns when we examined hospitals for a similar selection of codes, which implies that coding style is random

and specific to each hospital. We recommend that researchers who conduct this type of research consult with coders at all of the institutions included in their study to compile a list of feasible codes used for visits related to blood or body fluid exposure and attempt to understand coding practices.

ACKNOWLEDGMENTS

Financial support. R.C.M. was supported by a National Institutes of Health training grant through the Division of Infectious Diseases, Brown Medical School, the Miriam Hospital, from the National Institute on Drug Abuse (5 T32 DA13911). This study was supported in part by grants from the National Institutes of Health to the Brown/Lifespan/Tufts Centers for AIDS Research (P30 AI42853), the Rhode Island Foundation, and the Elizabeth Glaser Pediatric AIDS Foundation.

Potential conflicts of interest. All authors report no conflicts of interest relevant to this article. All authors submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest, and the conflicts that the editors consider relevant to this article are disclosed here.

Address correspondence to Roland C. Merchant, MD, MPH, ScD, Department of Emergency Medicine, Rhode Island Hospital, 593 Eddy Street, Claverick Building, Providence, RI 02903 (rmerchant@lifespan.org).

REFERENCES

- Varas-Lorenzo C, Castellsague J, Stang MR, Tomas L, Aguado J, Perez-Gutthann S. Positive predictive value of ICD-9 codes 410 and 411 in the identification of cases of acute coronary syndromes in the Saskatchewan Hospital Automated Database. *Pharmacoepidemiol Drug Saf* 2008;17:842–852.
- White RH, Garcia M, Sadeghi B, et al. Evaluation of the predictive value of ICD-9-CM coded administrative data for venous thromboembolism in the United States. *Thromb Res* 2010; 126(1):61–67.
- Semins MJ, Trock BJ, Matlaga BR. Validity of administrative coding in identifying patients with upper urinary tract calculi. *J Urol* 2010;184(1):190–192.
- Chibnik LB, Massarotti EM, Costenbader KH. Identification and validation of lupus nephritis cases using administrative data. *Lupus* 2010;19(6):741–743.
- Clark S, Gaeta TD, Kamarthi GS, Camargo CA. ICD-9 CM coding of emergency department visits for food and insect sting allergy. *Ann Epidemiol* 2006;16(9):696–700.
- Leibson CL, Naessens JM, Brown RD, Whisnant JP. Accuracy of hospital discharge abstracts for identifying stroke. *Stroke* 1994; 25(12):2348–2355.
- Chewning SJ, Nussman DS, Griffo ML, Kiebzak GM. Health care information processing: how accurate are the data? *J South Orthop Assoc* 1997;6(1):8–16.
- O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JE, Ashton CM. Measuring diagnoses: ICD code accuracy. *Health Serv Res* 2005;40(5):1620–1639.
- Farzandipour M, Sheikhtaheri A. Evaluation of factors influencing accuracy of principal procedure coding based on ICD-9-CM: an Iranian study. *Perspect Health Info Manag* 2009;6:5.
- Patkar NM, Curtis JR, Teng GG, et al. Administrative codes combined with medical records based criteria accurately identified bacterial infections among rheumatoid arthritis patients. *J Clin Epidemiol* 2009;62:321–327.
- Merchant RC, Becker BM, Mayer KH, Fuerch J, Schreck B. Emergency department blood or body fluid exposure evaluations and HIV postexposure prophylaxis usage. *Acad Emerg Med* 2003;10(12):1345–1353.
- Merchant RC, Mayer KH, Becker BM, Delong AK, Hogan JW. Predictors of the initiation of HIV postexposure prophylaxis in Rhode Island emergency departments. *AIDS Patient Care STDS* 2008;22(1):41–52.
- Merchant RC, Katzen JB, Mayer KH, Becker BM. Emergency department evaluations of non-percutaneous blood or body fluid exposures during cardiopulmonary resuscitation. *Prehosp Disaster Med* 2007;22(4):331–334.
- Merchant RC, Lau TC, Liu T, Mayer KH, Becker BM. Adult sexual assault evaluations at Rhode Island emergency departments, 1995–2001. *J Urban Health* 2008;86(1):43–53.
- Merchant RC, Nettleton JE, Mayer KH, Becker BM. Blood or body fluid exposures and HIV postexposure prophylaxis utilization among first responders. *Prehosp Emerg Care* 2009;13(1): 6–13.
- Merchant RC, Chee KJ, Liu T, Mayer KH. Incidence of visits for health care worker blood or body fluid exposures and HIV postexposure prophylaxis provision at Rhode Island emergency departments. *J Acquir Immune Defic Syndr* 2008;47(3):358–368.
- Merchant RC, Kelly ET, Mayer KH, Becker BM, Duffy SJ, Pughatch DL. Compliance in Rhode Island emergency departments with American Academy of Pediatrics recommendations for adolescent sexual assaults. *Pediatrics* 2008;121(6):1660–1667.
- Merchant RC, Nettleton JE, Mayer KH, Becker BM. HIV post-exposure prophylaxis among police and corrections officers. *Occup Med* 2008;58(7):502–505.
- Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1984.
- U.S. Census Bureau. *Metropolitan and micropolitan statistical areas*. November 2008. <http://www.census.gov/population/metro.html>.
- Stram DO, Lee JW. Variance components testing in the longitudinal mixed effects model. *Biometrics* 1994;50(4):1171–1177.
- Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis*. 6th ed. New York: Prentice Hall, 2007.
- Dingemans PM. ICD-9-CM classification coding in psychiatry. *J Clin Psychol* 1990;46(2):161–168.
- Throckmorton TW, Dunn W, Holmes T, Kuhn JT. Intraobserver and interobserver agreement of International Classification of Diseases, Ninth Revision codes in classifying shoulder instability. *J Shoulder Elbow Surg* 2009;18(2):199–203.