



Published in final edited form as:

South Med J. 2022 November ; 115(11): 801–805. doi:10.14423/SMJ.0000000000001463.

Trends in ICD-10-CM–Coded Administrative Datasets for Injury Surveillance and Research

Julia F. Costich, JD, PhD,

Dana B. Quesinberry, JD, DrPH,

Lara K. Daniels, MPH,

Ashley Bush, DrPH

Department of Health Management & Policy, Kentucky Injury Prevention & Research Center, College of Public Health, University of Kentucky, and the Kentucky Injury Prevention & Research Center, College of Public Health, University of Kentucky, Lexington.

Abstract

Objectives: Accurate injury surveillance depends on data quality in administrative datasets created for billing and reimbursement. Significant effort has been devoted to testing the ability of candidate injury case definitions to identify injury cases accurately in these datasets. We used interviews with experienced coders, informed by a review of the current literature, to identify three clinical coding trends that may affect the consistency of surveillance data: “clinical documentation improvement” (CDI),^A coding by treating clinicians, and certain electronic health record features.

Methods: An extensive literature review informed interviews with coding experts to identify potential issues in coding practice. To determine whether physician coding was associated with information loss, we analyzed data from two hospitals serving the same geographic area. One hospital had used physician coding of emergency department data for the past decade; the other used professional coders. We compared the proportion of emergency department records missing external cause of injury codes and assessed the variation for statistical significance.

Results: CDI audits review patient records to ensure that billing information includes every relevant *International Classification of Diseases, Tenth Revision, Clinical Modification* code. This

^AProduction Editor: Please check that this is the correct OA footnote for the level the author paid for.

^BL.K.D. and A.B. listed “government (not author’s own work)” in their COI statements. Was this the CDC funding or something else? Please clarify/specify.

^A“CDI” was defined different ways: “clinical documentation improvement” and “clinical documentation integrity.” Could you please update the definition for this abbreviation so it is consistent throughout?

This is an open access article distributed under the terms of the Creative Commons AttributionNon Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited.

Correspondence to Dr Julia F. Costich, Department of Health Management & Policy, University of Kentucky College of Public Health, 111 Washington Ave, Lexington, KY 40536. julia.costich@uky.edu.

The work cannot be changed in any way or used commercially without permission from the journal.A

Supplemental digital content is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal’s Web site (<http://sma.org/smj>).

The authors did not report any financial relationships or conflicts of interest.

A.B. has received compensation from Core State Violence Injury Prevention Program Surveillance Quality Improvement. BThe remaining authors did not report any financial relationships or conflicts of interest.

approach has increased payment rates awarded to Medicare Advantage plans because additional codes increase the patient acuity level and case mix index. The impact of CDI audits on injury data needs further investigation. The pilot analysis addressing information loss with physician coding found a higher level of external cause coding with clinician self-coding, possibly because of the coding software. Finally, widespread “copy and paste” in patient electronic health records has the potential to increase reported injuries.

Conclusions: Injury surveillance relies on billing and reimbursement records. Financial motivations may interfere with the consistency of surveillance findings and mislead injury epidemiologists. Further investigation is essential to ensure the integrity of surveillance findings.

Keywords

administrative data; data quality; ICD-10-CM

Injury epidemiologists and others who use administrative data for surveillance and research need a nuanced understanding of diagnostic coding practice.¹ Administrative datasets are coded with the *International Classification of Diseases, Tenth Revision, Clinical Modification* (ICD-10-CM).² These codes are intended to be used for billing and reimbursement rather than surveillance of specific diseases and conditions.³

Coders can only use diagnostic codes supported by clinical documentation in the patient record, but the scope of potential coding detail is broad. ICD-10-CM provides coding specialists with a wealth of options, most of which are not relevant to the administrative purposes for which clinical encounters are coded. As such, although ICD-10-CM-coded diagnostic information has the potential to provide injury epidemiologists with vastly more information than the ICD-9-CM, the codes in the records that populate the state administrative databases used by epidemiologists for injury surveillance generally lack this additional nuanced information. A 2019 study using a dataset of more than 60 million unique patients found that the use of a few of the more specific codes in the ICD-10-CM increased; however, changes varied markedly across diagnostic categories, and most new codes were not used.⁴

The practice of clinical coding requires extensive training and meticulous attention to detail. Because codes are used for billing and reimbursement, they are subject to both internal and external audits. In recent years, clinical documentation improvement (CDI) programs have been widely adopted to ensure that all relevant codes are captured in the billing documentation and that none of the codes diverge from the supporting clinical documentation. Proprietary algorithm-driven coding software assists the coder in navigating the tens of thousands of potential codes to ensure that billing reflects every eligible diagnosis.⁵

Some well-known phenomena may produce inaccuracy and confusion in the use by epidemiologists of ICD-10-CM injury codes for public health surveillance. For example, potential discrepancies may arise from aspects of coding audits, whereas another category of coding issues can emanate from features of electronic health record (EHR) systems, notably

their encouragement of code entry by clinical service providers. This study aimed to identify important sources of variation in coding and their potential impact on injury surveillance.

Methods

We conducted an extensive search of peer-reviewed and gray literature using PubMed, Google Scholar, and bibliographies from leading articles. The findings from these queries informed four semistructured telephone interviews with professional coders and auditors at two large health systems. The objectives of these interviews were identifying broad issues in coding practice rather than gleaning system-specific information. The interview questions can be found in Appendix A.

We also conducted a pilot quantitative analysis. One of the areas of concern raised by professional coders was the growing use of clinician coding, particularly in outpatient and emergency department (ED) settings. To test the hypothesis that provider coding led to a loss of information in coding, we used hospital administrative discharge data acquired from the Kentucky Office of Health Data and Analytics to identify two hospitals of similar size serving the same geographic area. Hospital A has used software-assisted coding by providers in the ED since before the inception of ICD-10-CM, whereas hospital B continues to use professional medical coders to code ED encounters. For the period 2016–2020, Kentucky resident injury-related ED encounters, excluding death, in both facilities were identified in accordance with guidance from the Centers for Disease Control and Prevention.⁶ External-cause-of-injury codes appear in sections V00 to Y99 of ICD-10-CM and are essential elements for the use of administrative claims data in injury surveillance. The percentage of external-cause-of-injury coding (hereafter “external cause coding”) was calculated using injury encounters with one or more external cause-of-injury or diagnosis codes, with external-cause-of-injury information as the numerator and the total number of injury encounters for each facility as the denominator. We compared the completeness of records using a two-proportion z test to determine the significance of the difference between the proportion of complete records in each hospital. Completeness was assessed using the following formula:

$$\frac{\text{Percentage of nonfatal injury ED visits with external-cause-of-injury coding}}{\text{All ED visit records with an injury diagnosis in any field}} = \frac{\text{All ED visit records with a code containing external-cause-of-injury information in any diagnosis or external-cause-of-injury field}}{\text{All ED visit records with an injury diagnosis in any field}}$$

The investigation reported in this study was approved by the University of Kentucky institutional review board.

Results

The key informants identified three areas of concern regarding ICD-10-CM coding. The first was clinical documentation improvement (also called clinical documentation integrity [CDI]). The threats to accurate interpretation that they identified in this area were upcoding

(the use of codes indicating a higher degree of severity than was actually present) and inclusion of codes related to past conditions that were not present on admission.

The second potential threat to accurate interpretation of ICD-10-CM codes in administrative claims data identified by the key informants arises from the tendency of EHRs to carry diagnoses forward after they have resolved, largely because of the “cut-and-paste” approach to clinical documentation.

The third threat identified was the increasing use of physician coding, particularly in outpatient and emergency settings, rather than employment of professional coders. This threat was the subject of our pilot quantitative investigation. When we compared clinician coding with that of coding specialists, our findings refuted the hypothesis that provider coding led to information loss.

We found significantly greater completeness ($P < 0.0001$) in external cause coding by clinicians than by professional coders, although both rates were high (95.55% vs 87.74%) (Table). Hospital A maintained higher coding completeness during the entire study period (Table); 95.6% of cases for hospital A included external cause codes, whereas this figure was 87.7% for hospital B. These preliminary findings suggest that the growing practice of clinician coding does not detract from the completeness of injury documentation in administrative datasets, but the topic would benefit from analysis encompassing a broader range of facilities.

Discussion

Clinical Documentation Improvement and Related Internal Audits

CDI auditing is ubiquitous in health system coding operations.⁷ CDI audits are typically conducted by nurses with special training in coding who review coded records for precision, accuracy, and completeness. CDI audits focus primarily on inpatient stays because they involve more patient care services and higher reimbursement than outpatient or ED visits. As such, the coding of inpatient stays is likely to be much more detailed and more accurate than outpatient or ED coding. Outpatient visits may be sampled for audit, and newly employed clinicians' outpatient records are more intensively audited until they become familiar with coding systems.

The Hierarchical Condition Category risk adjustment system has been used by the Centers for Medicare and Medicaid Services since 2006, ostensibly to assure Medicare Advantage (MA) plans that enrolling sicker patients will not put them at a financial disadvantage.⁸ As Gilfillan and Berwick⁹ explained, the original metrics for MA plans' assumption of risk were based on significant undercoding in the Medicare fee-for-service system. MA plans can increase their Medicare funding rate dramatically by identifying codes that, although valid, were not contemplated in the original Medicare risk adjustment methodology. The authors noted the following:

The starting point is to get as many diagnosis codes as possible. An entire industry has been created to do just that, leading to billion-dollar valuations for firms, like Signify Health, that

provide analytical tools to enable coding efforts or make home visits for plans and providers. Most plans now use artificial intelligence (AI) Hierarchical Condition Category tools to identify coding opportunities.

To the extent that the added diagnosis codes are included in surveillance data, the rate of specific conditions may be higher than historical trends would project.

CDI abuse has been the subject of several multimillion-dollar US Department of Justice settlements with health systems and MA plans in recent years. Typically, health systems use in-house or contracted reviewers to identify opportunities to bill for services at higher rates. In these cases, the individual who informs federal authorities of potentially illegal activity often is a member of the audit team who becomes concerned about ongoing pressure to exaggerate the severity of patients' diagnoses. Recent examples have involved MA plans that were using a contractor to review patient records for opportunities to document higher severity levels.^{10,11}

ED visits are the least likely to receive CDI attention or, indeed, any internal review for coding completeness and accuracy. The time pressures and brevity of ED care do not lend themselves to thoughtful coding under the best of circumstances, and when clinicians are doing their own coding in the ED, these problems are exacerbated. The interviewees noted a pervasive tendency toward general or "unspecified" code selections because time does not allow for a detailed perusal of more nuanced coding options. This finding has serious implications for using ED administrative databases for injury surveillance unless the patient is admitted to inpatient care from the ED.

EHR-Related Issues

EHRs have many benefits for clinicians, patients, and health systems, but they also raise some inherent barriers to accurate coding. One classic example is a function of the EHR problem list, which is used to identify the patient's current clinical issues. When a patient is seen repeatedly in the same facility over months or years, clinical problems may be cut and pasted from one encounter to the next regardless of whether they are actually current problems. A diagnosis that should be coded as "history of" an injury can be coded as if the injury were part of the presenting problem list. In this example, the EHR would erroneously add a case for the injury in question.²

Another feature of contemporary health care that discourages detailed coding is patient access to clinicians' notes. Patients could be offended by documentation of "morbid obesity," for example, as well as indicators of other stigmatized conditions such as drug or alcohol abuse, mental illness, or suicidal behavior.¹² Failure to document these conditions obviously has the potential to undermine the quality of care as well as accurate diagnostic reporting. With regard to injury surveillance, the external cause of injury or comorbidities (eg, associated with intoxication or physical abuse) may be more sensitive than the injury diagnosis itself.

Physician Coding

Discharge diagnosis coding practices and quality may not be consistent across hospitals and hospital staff. Coding decisions may at times be influenced by billing.¹³ EHR platforms encourage coding by clinicians themselves in outpatient settings. In our interviews, the professional coders voiced deep concern about this trend because clinical service providers are not typically trained in coding. Clinicians' time is their most important asset, and they are not motivated to include detailed coding elements. The EHR software generates a series of codes representing the diagnoses most frequently seen in the clinician's patient population, and the clinician will select diagnoses primarily if not exclusively from that list. Nuanced additions regarding the nature and circumstances of a patient's injury do not affect reimbursement for the patient care encounter. Health systems have limited capacity to audit outpatient encounters, and the need for rapid billing submission puts pressure on the clinical system to limit coding to the minimum necessary to support charges. The current trend toward clinician coding thus may threaten the usefulness of outpatient datasets for injury surveillance.

Interviews with professional coders identified three potential sources of variation in ICD-10-CM coding that could affect the accuracy and consistency of injury surveillance using ED records. First, features of EHRs and their interaction with user practice may perpetuate injury diagnoses in the documentation of patient encounters of care for which the injury is no longer under treatment. Second, clinicians are increasingly doing their own coding and may have coding practices that vary from coding specialists. Most notably, the growing use of clinical documentation improvement/integrity reviews, particularly for patients in MA plans, may increase the number and severity of listed ICD-10-CM codes. If this practice adds a significant number of injury codes, then it will give the impression that the prevalence of injury has increased in the millions of MA beneficiaries. All of these issues would benefit from more extensive investigation; however, CDI reviews focus on inpatient hospitalizations, for which injury surveillance is typically limited to the principal diagnosis, so the impact of CDI in the ED context may be limited.

Billing for clinical services is driven primarily by the Centers for Medicare and Medicaid Services diagnosis-related group system, which includes 740 discrete classifications.¹⁴ In contrast, ICD-10-CM includes 72,616 codes as of 2021—nearly a 100-fold difference. Although the vast repertoire available in ICD-10-CM has much to offer the field of injury surveillance, it is largely irrelevant to the clinical practitioner or coder and appears to be ignored, particularly in outpatient and ED settings. As one coding expert noted, coding in these settings is “numbers chasing money,” and the more rapid the chase, the faster the revenue cycle turns visits into reimbursement.

The field of injury epidemiology can use three strategies to reconcile the grand scheme of ICD-10-CM coding with the reality of its use by coding professionals and clinical practitioners: match expectations to reality, explore alternative information sources, and pursue continual quality improvement.

Match Expectations to Reality

Epidemiologists should ask what data elements are essential for the statistical grounding of public health policy. Conversations about coding protocols and algorithms are more likely to succeed if they acknowledge the constraints and incentives of coding practice.

Explore Alternative Information Sources

Although ICD-10-CM coding practice requires strict correlation between clinical documentation and code assignment, epidemiological inquiry is not so limited. Additional documentation in the patient's medical record can yield insights beyond the codes themselves. Administrative datasets include procedure codes and information about patient transfers that may suggest factors of interest.

Pursue Surveillance Quality Improvement

Matching case definitions to clinical documentation via ICD-10-CM codes requires meticulous record reviews such as those conducted during the 5 years of the Centers for Disease Control and Prevention's Surveillance Quality Improvement initiative.¹⁵⁻¹⁸

Limitations

Our analysis is based on the expert opinion of coding specialists with decades of experience, supplemented by a review of the emerging literature on coding issues in ICD-10-CM and an analysis of data from two hospitals. Findings may not be generalizable to the much broader range of facilities and clinicians across the United States. The subject matter appears to be evolving as government auditors and the US Department of Justice identify the potential magnitude of the cost increases associated with changes in coding practice. Thus, our findings are limited to a point in time and would benefit from longitudinal assessments.

Conclusions

Machine learning and other AI-based methods may ultimately overcome the dependence of injury surveillance on administrative claims data, but their implementation will require investments far exceeding those currently available to state public health agencies. Investigators must be aware of systematic sources of divergence between the diagnostic information that administrative data appear to convey and the clinical records supporting them. Until AI development moves public health surveillance beyond reliance on administrative claims data, continued support for surveillance quality improvement is essential to ensure the integrity of injury epidemiology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work would not have been possible without the wealth of information contributed by the professional coder interviewees and their generosity in sharing their time and expertise. Svetla Slavova, PhD, and Renee Johnson, RPT, MSPH, provided invaluable guidance. Thanks to the University of Kentucky College of Public Health's Office of Scientific Writing for extensive assistance.

This study was funded by award number 5NU17CE924846–05 from the Centers for Disease Control and Prevention under the Surveillance Quality Improvement initiative. All of the findings are those of the authors and do not reflect the influence or positions of the Centers for Disease Control and Prevention.

References

1. Johnson RL, Hedegaard H, Pasalic ES, et al. Use of ICD-10-CM coded hospitalisation and emergency department data for injury surveillance. *Inj Prev* 2021;27:i1–i2. [PubMed: 33674325]
2. World Health Organization. International Statistical Classification of Diseases and Related Health Problems (ICD). <https://www.who.int/classifications/classification-of-diseases>. Accessed August 2021.
3. Fung KW, Xu J, Bodenreider O. The new International Classification of Diseases 11th edition: a comparative analysis with ICD-10 and ICD-10-CM. *J Am Med Inform Assoc* 2020;27:738–746. [PubMed: 32364236]
4. Sivashankaran S, Borsi JP, Yoho A. Have ICD-10 coding practices changed since 2015? *AMIA Annu Symp Proc* 2019;2019:804–811. [PubMed: 32308876]
5. American Health Information Management Association. Computer-assisted coding toolkit. https://library.ahima.org/doc?oid=300442#.YWLis_kpCuk. Accessed October 2021.
6. Thomas KE, Johnson RL. State injury indicators report: instructions for preparing 2019 data. https://www.cdc.gov/injury/pdfs/2019_state_injury_indicator_instructions-508.pdf. Accessed August 2021.
7. Ericson C, Pratt RL, Yeun AP. How do you distinguish clinical documentation: *J Am Health Inform Manag Assoc*. <https://journal.ahima.org/how-do-you-distinguish-clinical-documentation>. Accessed September 2021.^B
8. Tozzi J. The hunt for the riskiest, most lucrative patients. <https://www.bloomberg.com/news/articles/2021-10-13/private-medicare-advantage-programs-don-t-save-government-money-critics-say>. Published October 13, 2021. Accessed August 28, 2022.
9. Gilfillan R, Berwick D. Medicare Advantage, direct contracting, and the Medicare “money machine.” Part 1: the risk-score game. https://www.healthaffairs.org/doi/10.1377/hblog20210927.6239/full/?utm_medium=email&utm_source=hasu&utm_source=Newsletter&utm_campaign=blog&utm_content=gilfillan&utm_content=Low-Density+Residential+Zoning%2C+Banning+Surprise+Bills+%26+More&vgo_ee=WTegeLuWfK8jamDctZPMI%3D%3D. Published September 29, 2021. Accessed October 2021.
10. US Department of Justice. United States intervenes and files complaint in false claims act suit against health insurer for submitting unsupported diagnoses to the Medicare Advantage program. <https://www.justice.gov/opa/pr/united-states-intervenes-and-files-complaint-false-claims-act-suit-against-health-insurer>. Published September 14, 2021. Accessed September 2021.
11. US Department of Justice. Government intervenes in false claims act lawsuits against Kaiser Permanente affiliates for submitting inaccurate diagnosis codes to the Medicare Advantage program. Available: <https://www.justice.gov/opa/pr/government-intervenes-false-claims-act-lawsuits-against-kaiser-permanente-affiliates>. Published July 30, 2021. Accessed October 2021.
12. DesRoches CM, Leveille S, Bell SK, et al. The views and experiences of clinicians sharing medical record notes with patients. *JAMA Netw Open* 2020;3:e201753.
13. Council of State and Territorial Epidemiologists. Nonfatal opioid overdose standardized surveillance case definition. Interim Cross-Cutting Committee Report-19. https://cdn.ymaws.com/www.cste.org/resource/resmgr/ps/2019ps/PS_Nonfatal_Opioid_Overdose_.pdf. Revised 2018. Accessed September 2021.
14. Centers for Medicare and Medicaid Services. MS-DRG classifications and software. <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/MS-DRG-Classifications-and-Software>. Accessed September 2021.

^BThe URL for reference 7 no longer works. Please replace this with a working URL that readers can access.

^BThe URL for reference 7 no longer works. Please replace this with a working URL that readers can access.

15. Gabella BA, Hathaway JE, Hume B, et al. Multisite medical record review of emergency department visits for traumatic brain injury. *Inj Prev* 2021;27:i42–i48. [PubMed: 33674332]
16. Peterson A, Gabella BA, Johnson J, et al. Multisite medical record review of emergency department visits for unspecified injury of head following the ICD-10-CM coding transition. *Inj Prev* 2021;27:i13–i18. [PubMed: 33674328]
17. Hansen A, Slavova D, Cooper G, et al. An emergency department medical record review for adolescent intentional self-harm injuries. *Inj Epidemiol* 2021;8:3. [PubMed: 33413622]
18. Warwick J, Slavova S, Bush J, et al. Validation of ICD-10-CM surveillance codes for traumatic brain injury inpatient hospitalizations. *Brain Inj* 2020;34:1763–1770. [PubMed: 33280404]

Key Points

- The practice known as clinical data improvement has the goal of capturing every *International Classification of Diseases, Tenth Revision, Clinical Modification* code for each patient encounter, and thus may increase recorded injury codes.
- Clinical service providers are increasingly required to enter algorithm-driven codes for the work they have performed, and this coding practice may alter the mix of injury codes in administrative datasets.
- The use of electronic health records has well-known features that may carry injury codes forward into patient care encounters in which injuries are no longer under treatment.

Frequency and percentage of injury-related ED encounters with external-cause-of-injury coding in two Kentucky hospitals, 2016–2020

Table.

Hospital	Injury encounters with an external-cause-of-injury code, n	Total injury encounters, N	Completeness for the study period, %	Average monthly completeness, %	Lowest month completeness, %	Highest month completeness, %	P
A	23,762	24,868	95.55	95.43	84.95	99.74	<0.0001 ^a
B	42,005	47,874	87.74	87.74	80.39	94.74	

ED, emergency department.

^aThe P value is the result of the testing of differences in column labeled “Completeness for the study period, %.”