8-26-2022

Accuracy of ICD-10-CM encounter diagnoses from health records for identifying self-harm events

Gregory E. Simon

Susan M. Shortreed

Jennifer M. Boggs

Gregory N. Clarke

Rebecca C. Rossom

See next page for additional authors

Recommended Citation

Simon GE, Shortreed SM, Boggs JM, Clarke GN, Rossom RC, Richards JE, Beck A, Ahmedani BK, Coleman KJ, Bhakta B, Stewart CC, Sterling S, Schoenbaum M, Coley RY, Stone M, Mosholder AD, and Yaseen ZS. Accuracy of ICD-10-CM encounter diagnoses from health records for identifying self-harm events. J Am Med Inform Assoc 2022.

Authors

Gregory E. Simon, Susan M. Shortreed, Jennifer M. Boggs, Gregory N. Clarke, Rebecca C. Rossom, Julie E. Richards, Arne Beck, Brian K. Ahmedani, Karen J. Coleman, Bhumi Bhakta, Christine C. Stewart, Stacy Sterling, Michael Schoenbaum, R. Yates Coley, Marc Stone, Andrew D. Mosholder, and Zimri S. Yaseen

Research and Applications

Accuracy of ICD-10-CM encounter diagnoses from health records for identifying self-harm events

Gregory E. Simon[1], Susan M. Shortreed[1], Jennifer M. Boggs[2], Gregory N. Clarke[3], Rebecca C. Rossom [4], Julie E. Richards[1], Arne Beck[2], Brian K. Ahmedani[5], Karen J. Coleman[6], Bhumi Bhakta[6], Christine C. Stewart[1], Stacy Sterling[7], Michael Schoenbaum[8], R. Yates Coley[1], Marc Stone[9], Andrew D. Mosholder[9], and Zimri S. Yaseen [9]

[1]Kaiser Permanente Washington Health Research Institute, Seattle, Washington, USA, [2]Kaiser Permanente Colorado Institute for Health Research, Denver, Colorado, USA, [3]Kaiser Permanente Northwest Center for Health Research, Portland, Oregon, USA, [4]HealthPartners Institute, Minneapolis, Minnesota, USA, [5]Center for Health Policy and Services Research, Henry Ford Health, Detroit, Michigan, USA, [6]Kaiser Permanente Southern California Department of Research and Evaluation, Pasadena, California, USA, [7]Kaiser Permanente Northern California Division of Research, Oakland, California, USA, [8]National Institute of Mental Health, Bethesda, Maryland, USA, and [9]U.S. Food and Drug Administration, Silver Spring, Maryland, USA

Corresponding Author: Gregory E. Simon, MD, MPH, Kaiser Permanente Washington Health Research Institute, 1730 Minor Ave #1600, Seattle, WA 98101, USA; gregory.e.simon@kp.org

ABSTRACT

Objective: Assess the accuracy of ICD-10-CM coding of self-harm injuries and poisonings to identify self-harm events.

Materials and Methods: In 7 integrated health systems, records data identified patients reporting frequent suicidal ideation. Records then identified subsequent ICD-10-CM injury and poisoning codes indicating self-harm as well as selected codes in 3 categories where uncoded self-harm events might be found: injuries and poisonings coded as undetermined intent, those coded accidental, and injuries with no coding of intent. For injury and poisoning encounters with diagnoses in those 4 groups, relevant clinical text was extracted from records and assessed by a blinded panel regarding documentation of self-harm intent.

Results: Diagnostic codes selected for review include all codes for self-harm, 43 codes for undetermined intent, 26 codes for accidental intent, and 46 codes for injuries without coding of intent. Clinical text was available for review for 285 events originally coded as self-harm, 85 coded as undetermined intent, 302 coded as accidents, and 438 injury events with no coding of intent. Blinded review of full-text clinical records found documentation of self-harm intent in 254 (89.1%) of those originally coded as self-harm, 24 (28.2%) of those coded as undetermined, 24 (7.9%) of those coded as accidental, and 48 (11.0%) of those without coding of intent.

Conclusions: Among patients at high risk, nearly 90% of injuries and poisonings with ICD-10-CM coding of self-harm have documentation of self-harm intent. Reliance on ICD-10-CM coding of intent to identify self-harm would fail to include a small proportion of true self-harm events.

Key words: suicide, self-harm, ICD-10-CM, diagnoses, coding

INTRODUCTION

Public health surveillance, quality improvement, and epidemiologic research regarding suicidal behavior often rely on health system records to identify self-harm events. Population-level surveillance of self-harm depends on insurance claims from emergency department and inpatient treatment settings.[1] National suicide prevention efforts recommend use of records data to monitor progress as a core improvement strategy.[2] Evaluations of clinical interventions, including medications or psychosocial interventions, and care improvement efforts also rely on diagnoses in health records to evaluate clinical impact.[3–5] Accuracy of those coded encounter diagnoses depends on both the coding system in place and implementation of that system in practice.

Previous research regarding accuracy of encounter diagnoses to identify self-harm has yielded mixed results regarding both over- and under-ascertainment.[6,7] Under ICD-9-CM,[8] coding of intent (ie, accident, assault, self-harm) required separate cause-of-injury codes or E-codes. We have previously reported that recording of cause-of-injury codes varied widely across health systems and within health systems over time.[9] Among studies reviewing clinical notes from encounters with ICD-9-CM E-codes for self-inflicted injury or poisoning,[10–12] rates of confirmation for intentional self-harm ranged between 36% and 100%. Assessments of how often ICD-9-CM E-codes failed to identify self-harm have used varying methods and have reported widely varying estimates of the proportion of self-harm events not captured.[6,7,13] We have previously reported that clinical notes for injuries and poisonings receiving ICD-9-CM E-codes for "undetermined intent" often included clear documentation of self-harm.[14] These disparate findings regarding accuracy of coding likely reflect both true differences between settings in coding practices and differences in methods used to assess accuracy. Varying findings led to varying recommendations regarding use of ICD-9-CM diagnostic codes to identify self-harm.[6,7,14]

Classification of injuries and poisonings changed significantly with the transition from ICD-9-CM to ICD-10-CM in October 2015. Under ICD-10-CM,[15] coding of intent is integrated into primary codes for all poisonings and some injuries, so that separate cause-of-injury codes are no longer required. The ICD-10-CM system also requires more detailed specification of injury and poisoning categories. We previously reported that this transition led to a marked decrease in injuries and poisonings coded as undetermined intent and a corresponding increase in coding of self-harm.[16] We are aware of no research assessing accuracy of self-harm coding since the transition to ICD-10-CM.

Here, we describe a systematic assessment of the accuracy of ICD-10-CM coding to identify self-harm events among patients at increased risk for suicidal behavior. We use clinical notes to estimate the proportion of injuries or poisonings coded as self-harm for which full-text records confirm self-harm intent, as well as the proportions of selected injuries or poisonings not coded as self-harm for which full-text records do indicate self-harm intent. Regarding events not coded as self-harm, we selected 3 categories of codes in which missing self-harm events might be found: injuries and poisonings coded as having undetermined intent, those coded as accidents, and injuries without coding of intent. Findings regarding accuracy of ICD-10-CM coding for self-harm should be useful for health systems evaluating suicide prevention efforts and researchers using records data to identify self-harm events.

MATERIALS AND METHODS

Settings

This research was conducted in 7 integrated health systems participating in the Mental Health Research Network.[17] Responsible institutional review boards reviewed and approved use of records data in this research. All study sites extract and translate health system coded electronic health records (EHRs) data and insurance claims data into compatible research data warehouses following the Health Care Systems Research Network Virtual Data Warehouse model.[18] Clinical text is available in health system EHR databases for all encounters at health system facilities and a portion of external encounters, but some external encounters captured by insurance claims are not recorded in EHRs.

Summary of design

An initial wave of chart reviews included 4 health systems (Health-Partners and the Colorado, Northwest, and Washington regions of Kaiser Permanente) participating in a large pragmatic trial of outreach programs to prevent suicide attempt or other self-harm.[3,19] That trial included patients aged 18 or older enrolled in the health system who completed a Patient Health Questionnaire or PHQ-9 depression questionnaire[20,21] between October 2015 and September 2018 and reported thoughts of suicide or self-harm (ie, the ninth question of the PHQ-9) either "more than half the days" or "nearly every day." This initial wave of reviews focused on potentially missed self-harm events among injury and poisoning events NOT coded as self-harm. Reviews considered outpatient, emergency department, or inpatient encounters among trial participants over 18 months after randomization.

A second wave of reviews added 3 health systems: Henry Ford Health and the Northern California and Southern California Kaiser Permanente regions. The samples in these 3 additional health systems paralleled the pragmatic trial sample: adult outpatients who completed a PHQ-9 depression questionnaire between October 2015 and September 2018 and reported frequent thoughts of suicide or self-harm. This second wave examined potentially missed self-harm events in the 3 additional health systems and also examined records from all 7 health systems for documentation of self-harm among injuries and poisonings that were originally coded as self-harm. Reviews considered outpatient, emergency department, and inpatient encounters over 18 months after completion of a PHQ-9 questionnaire.

As shown in Figure 1, evaluating the accuracy of specific diagnostic codes for identifying self-harm included 5 steps: identifying groups of relevant injury and poisoning codes, selecting specific codes where missed self-harm events might be found, using those specific codes to identify injury and poisoning events for review of clinical text, abstraction of relevant clinical text for those identified events, and blinded grading of that clinical text for documentation of self-harm intent.

Identification of relevant coding categories

ICD-10-CM codes indicating self-harm included all injury codes in the range of X71 through X83 as well as all codes for poisoning (T36 through T65), unclassified injury (T14), and asphyxiation (T71) that included a specifier for self-harm intent. Codes where missed self-harm events might occur included 3 categories of ICD-10-CM injury and poisoning codes:
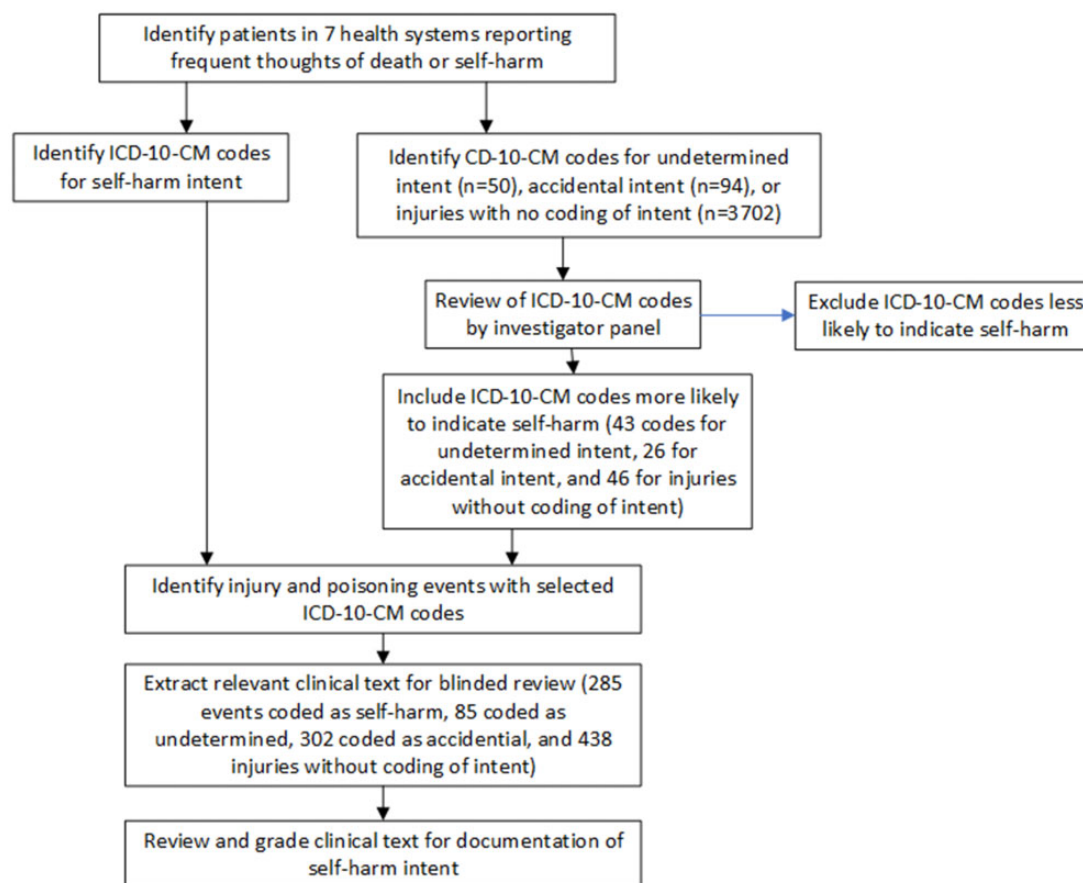
Figure 1. Study flow chart.

- *Codes indicating undetermined intent* included all injury codes in the range of Y21 through Y33 as well as codes for poisoning (T36 through T65), injury (T14), and asphyxiation (T71) that included a specifier for undetermined intent.
- *Codes indicating accidental intent* included codes for poisoning (T36 through T65), injury (T14), and asphyxiation (T71) that included a specifier for accidental intent.
- *Injury diagnoses without coding of intent* included codes in the range from S00 through T32 that were not accompanied by an external cause code in the range from V00 through Y99.

Selection of codes where uncoded self-harm events might be found

Given the large numbers of codes and potentially large numbers of events not originally coded as self-harm, a panel of 5 investigators (GES, RCR, AB, GNC, and JMB) reviewed all codes in these groups that were actually observed in the study sample to identify specific codes where missed self-harm events might be more likely. This review considered only code descriptors (eg, poisoning by unspecified narcotic, accidental intent) from the version of ICD-10-CM in use at the time the diagnosis was recorded and did not consider any other characteristics of any individual event. Among codes for undetermined intent, panel members were asked to identify those unlikely to represent self-harm. Codes identified as unlikely by at least 3 panel members were excluded from record review, with all other codes included. Among codes for accidental intent or injuries without coding of intent, panel members were asked to identify codes in-

dicating common or likely mechanisms of self-harm. Codes identified as likely by at least 3 panel members were included in record review, with all other codes excluded.

Selection of injury and poisoning events for review

The resulting 4 lists of ICD-10-CM codes (all codes indicating self-harm and selected codes for undetermined intent, accidental intent, or injuries without coding of intent) were then used to identify injury or poisoning events eligible for review of clinical text. The first wave of chart reviews included all eligible events originally assigned selected codes for undetermined intent, accidental intent, or without coding of intent in the pragmatic trial sample. The second wave included randomly selected events in those same 3 categories in the 3 additional health systems (up to 50 in each coding group from each health system) as well as randomly selected events originally coded as self-harm from all 7 health systems (up to 50 from each health system).

Review-eligible events were defined by the first occurrence of at least one reviewable diagnosis code as described above. An event with any code indicating self-harm was placed in that group regardless of other codes (accident, undetermined intent, no coding of intent) assigned. A single injury or poisoning event not coded as self-harm could be selected into more than one of the other groups. For example, a selected code for poisoning of undetermined intent and a selected code for accidental poisoning for the same event would result in selection into both groups.

Extraction of clinical text

For each event, chart abstractors were instructed to consider clinical notes from any encounters (outpatient, emergency department, inpatient, and telephone encounters) within 14 days before or after the date of the qualifying diagnosis. Abstraction typically began with any encounter on the diagnosis date, extending to encounters before and after that date until clear documentation of intent was identified or until all encounters during the interval were reviewed. Abstractors identified and extracted text during the ±14-day period that was most relevant to the intent of the injury or poisoning selected for review, including text from nursing notes, treating clinicians' notes, and direct quotes from patients. Abstractors were advised to specifically identify text that would clarify presence or absence of self-harm intent, including both suicidal intent and intentional self-harm not necessarily accompanied by intent to die (ie, non-suicidal self-injury). Abstractors redacted any information that might describe past injuries or poisonings, describe injuries or poisonings after the index date, or allow re-identification of individual patients or healthcare providers.

Blinded grading of clinical text

All text extracted for each event was then presented to a panel of 6 study investigators (graders), with each event considered by 3 graders and each grader considering approximately half of all events. Graders were blinded to original coding of self-harm and to ratings of other graders. Graders were advised that the sample included a mixture of events originally coded as self-harm and events not coded as self-harm. Graders were instructed to grade text from each event as indicating self-harm intent or not (a forced-choice vote of yes or no) and to separately grade confidence in that forced-choice classification (high, medium, or low). Graders were instructed to assess documentation of self-harm intent regardless of intent to die or potential lethality of the injury or poisoning.

Data analysis

Descriptive analyses examined the distribution of graders' self-harm votes and confidence ratings across the 4 groups defined by original coding. Simple analyses considered only the majority of the 3 graders' votes for each event (ie, 2 or 3 votes for self-harm intent = self-harm documented, 2 or 3 votes against self-harm intent = self-harm not documented). To also account for confidence ratings, weighted analyses assigned a score to each grading of each event where a vote for self-harm intent carried a weight ranging from +1 (yes vote with low confidence) to +3 (yes vote with high confidence) and a vote against self-harm intent carried a weight ranging from −1 (no vote with low confidence) to −3 (no vote with high confidence). Summing across 3 graders, the summary score for each event ranged from −9 (3 votes against self-harm intent high confidence) to +9 (3 votes for self-harm intent with high confidence). Secondary analyses examined consistency of findings across the 7 health systems in each of the 4 coding groups, using the weighted method described above. Exact confidence limits for proportions were calculated by the Clopper–Pearson method,[22] agreement among raters was evaluated using Fleiss' kappa statistic,[23] and between-group comparisons of proportions were evaluated using Fisher's exact test.[24] Analyses were performed using SPSS version 22, except for calculation of Fleiss' kappa scores performed using Stata version 15.1.

RESULTS

Selection of codes where uncoded self-harm events might be found

Injury and poisoning codes recorded in the study sample during the follow-up period included 50 ICD-10-CM codes for injuries or poisonings with undetermined intent, 94 codes for injuries or poisonings with accidental intent, and 3702 codes for injuries not accompanied by coding of intent. Among the 50 undetermined intent codes, the panel of investigators excluded 7 codes as unlikely mechanisms of self-harm (eg, spider bite), leaving 43 codes (86% of codes representing 90% of undetermined intent events) as eligible for review. Among 94 accidental injury/poisoning codes, the panel of investigators identified 26 (28% of codes representing 19% of accidental intent events) as common mechanisms of self-harm to be included in review. Among 3702 injury codes with no coding of intent, the panel of investigators identified 46 (1% of codes representing fewer than 1% of injury events without coding of intent) as common mechanisms of self-harm to be included in review. The most frequent included and excluded codes in each group are shown in Table 1, and a complete list of included codes is provided in Supplementary Appendix A.

Selection of injury and poisoning events for review and extraction of clinical text

The code lists and selection procedures described above identified the following numbers of events for review across the 7 health systems: 304 events coded as self-harm, 97 coded as having undetermined intent, 348 coded as having accidental intent, and 466 injuries without coding of intent. Review of all clinical notes within 14 days before or after each of those events found no relevant text (ie, no encounters in the EHR with any mention of injury or poisoning) in 19 (6%) of events coded as self-harm, 12 (12%) of undetermined intent events, 46 (13%) of accidental intent events, and 28 (6%) of injury events without coding of intent. Clinical notes could be missing for encounters occurring outside the health system but identified by insurance claims. Exclusion of those events with missing clinical notes left 285 events coded as self-harm, 85 events receiving a code for undetermined intent, 302 events receiving a code for accidental intent, and 438 injury events with no coding of intent.

Blinded grading of clinical text

Examples of extracted text with self-harm ratings and confidence scores are shown in Table 2. Table 3 displays agreement among raters and the proportions of events found to have documentation of self-harm intent using different scoring methods and confidence thresholds. Fleiss kappa statistics for agreement of yes/no classification among 3 graders indicated very good agreement for accidents with no original coding of intent, good agreement for events originally coded as self-harm, and moderate agreement for events originally coded as having undetermined or accidental intent. Following a simple majority rule, the proportion of events judged to have documentation of self-harm intent ranged from 7.0% among those originally coded as accidental to 87.7% among those originally coded as self-harm. Distributions of weighted summary scores for the 4 coding groups are shown in Figure 2. For events originally coded as self-harm, approximately 80% of summary scores indicated unanimous votes for self-harm intent with moderate or high confidence (ie, scores of +6 or higher), but approximately 6% indicated unanimous votes against self-harm intent with moderate or high confidence (ie, scores of −6 or lower). Injuries without original coding of intent

Table 1. Most common injury and poisoning diagnosis codes selected for inclusion and exclusion

	Included in record review		Excluded from record review	
Undetermined intent	T50.904A—Poisoning by unspecified drug, undetermined intent	38%	T63.304A—Toxic effect of spider venom, undetermined intent	2%
	T42.4X4A—Poisoning by benzodiazepine, undetermined intent	5%	T63.444A—Toxic effect of bee venom, undetermined intent	2%
	T65.94XA—Toxic effect of unspecified substance, undetermined intent	4%	T63.464A—Toxic effect of wasp venom, undetermined intent	2%
	T51.94XA—Toxic effect of unspecified alcohol, undetermined intent	4%	T59.3X4A—Toxic effect of lacrimogenic gas, undetermined intent	1%
	T43.594A—Poisoning by antipsychotics, undetermined intent	4%	T63.484A—Toxic effect of other arthropod venom, undetermined intent	1%
Accidental	T42.4X1A—Poisoning by benzodiazepine, accidental	5%	T50.901A—Poisoning by unspecified drug, accidental	21%
	T43.591A—Poisoning by antipsychotic, accidental	4%	T63.441A—Toxic effect of bee venom, accidental	10%
	T40.2X1A—Poisoning by other opioid, accidental	3%	T63.481A—Toxic effect of other arthropod venom, accidental	9%
	T40.601A—Poisoning by unspecified narcotic, accidental	3%	T56.891A—Toxic effect of other metals, accidental	8%
	T42.6X1A—Poisoning by antiepileptic or sedative/hypnotic, accidental	3%	T63.461A—Toxic effect of wasp venom, accidental	2%
No coding of intent	S51.812A—Laceration without foreign body of left forearm	<1%	S39.012A—Strain of muscle, fascia, and tendon of lower back	7%
	S51.811A—Laceration without foreign body of right forearm	<1%	S16.1XXA—Strain of muscle, fascia, and tendon of neck	5%
	S61.512A—Laceration without foreign body of left wrist	<1%	S09.90XA—Unspecified injury of head	3%
	S61.511A—Laceration without foreign body of right wrist	<1%	T14.8XXA—Other injury of unspecified body region	1%
	S51.802A—Unspecified open wound of left forearm	<1%	S93.401A—Sprain of unspecified ligament of right ankle	<1%

Note: Percentages indicate proportion of events in that coding group receiving that individual code (eg, 38% of all events with undetermined intent codes received code T50.904A).

showed the opposite pattern: over 80% with scores indicating unanimous votes against self-harm intent with moderate or high confidence and approximately 7% indicating unanimous votes for self-harm intent with moderate or high confidence. Events originally coded as accidental had over 60% of events indicating unanimous votes against self-harm intent with moderate or high confidence, but also had approximately 25% indicating uncertainty or disagreement regarding presence or absence of self-harm intent (ie, summary scores in the range from −3 to +3). Events originally coded as having undetermined intent showed the greatest uncertainty or disagreement, with summary scores distributed throughout the range and over half in the range from −3 to +3. As shown in Supplementary Appendix C, summary scores in the range from −3 to +3 more often represented consistent uncertainty (low confidence from all raters) than clear disagreement (confident ratings in opposite directions). As shown in Table 3, using a weighted summary score threshold of +1 or higher yielded very similar results to the simple majority rule, with rates of documented self-harm intent ranging from 7.9% for those originally coded as accidental to 89.1% for those originally coded as self-harm. For the 2 categories with clear original coding of intent, those originally coded as self-harm or accidental, a stricter summary score threshold might be appropriate to re-classify or override the original coding (right column of Table 3). For events originally coded as accidental, requiring a summary score of +3 or higher (a threshold equivalent to 2 votes for self-harm intent with moderate confidence and 1 vote against self-harm intent with low confidence) would lead to 5.6% of events re-classified as having

documentation of self-harm intent. For events originally coded as self-harm, requiring a summary score of −3 or lower (a threshold equivalent to 2 votes against self-harm intent with moderate confidence and 1 vote for self-harm intent with low confidence) would lead to 91.6% of events remaining classified as self-harm and 8.4% re-classified as not having self-harm intent.

Figure 3 shows variation in results across the 7 health systems for each of the 4 coding groups. For events originally coded as self-harm, undetermined intent, or accidental intent, the proportion of events judged to have some documentation of self-harm intent (summary score > 0) did not differ significantly across health systems. For injuries without coding of intent, the proportion judged to have documentation of self-harm intent varied from 2% to 30% across health systems (exact $P < .001$).

DISCUSSION

In this first systematic assessment of self-harm coding under ICD-10-CM, findings are generally reassuring regarding use of intent coding to identify self-harm events and suggest improvement in coding accuracy after the transition from ICD-9-CM to ICD-10-CM. Nearly 90% of reviewable events coded as self-harm had documentation of self-harm intent in clinical notes. In 3 groups of events not coded as self-harm but selected because of codes more likely to represent missed self-harm, the proportions with documentation of self-harm intent ranged from approximately 8% (among injuries and

Table 2. Examples of abstracted text and grader evaluations

Relevant text extracted from clinical notes	Self-harm ratings			Confidence ratings		
Presents to UC post-cutting on her wrists	Yes	Yes	Yes	High	Medium	Medium
She was assaulted.	No	No	No	High	High	High
Presents unresponsive in laboratory. She was just seen in in clinic for planned detox medical clearance. She had just come to lab as part of medical clearance assessment. Her partner says she drank wine this morning, estimates around 8 ounces. She has access to Ativan. She does not have access to opiates. She had ETOH withdraw seizure last week per PC physician who was evaluating her for detox today.	No	No	No	Medium	Low	High
Pt brought to [hospital] after motor vehicle accident where pt sustained closed sternum fracture. Regarding his car accident… He took his father's car to [Location] to gamble and drink, and he reports having no memory of getting in his car and driving or the accident. He tells me he does not know if it was a suicide attempt.	No	No	No	Low	Low	Low

Table 3. Proportion of injury or poisoning events found to have documentation of self-harm intent following blinded review of full-text clinical records using different thresholds

Original coding	Sampled N[a]	Fleiss' Kappa	Majority yes votes		Weighted summary score >0		Stricter rule to re-classify	
			N (%)	95% CI	N (%)	95% CI	N (%)	95% CI
Self-harm	285	0.68	250 (87.7%)	83.3%–91.3%	254 (89.1%)	84.9%–92.5%	261 (91.6%)	87.7%–94.5%
Undetermined intent	85	0.60	27 (31.8%)	22.1%–42.8%	24 (28.2%)	19.0%–39.0%		
Accidental intent	302	0.45	21 (7.0%)	4.4%–10.5%	24 (7.9%)	5.2%–11.6%	17 (5.6%)	3.3%–8.9%
No coding of intent	438	0.85	49 (11.2%)	8.4%–14.5%	48 (11.0%)	8.3%–14.4%		

[a]Note that the numbers sampled in each stratum are not proportional to representation in the population, so results cannot be summed across strata. Furthermore, the relative sizes of those strata will differ across samples, depending on the risk-level in the population and coding practices in the setting.

poisonings originally coded as accidents) to approximately 30% (among those coded as having undetermined intent).

In this sample of reviewed events, ICD-10-CM coding of self-harm intent had positive predictive value or precision approaching 90% when judged by documentation of self-harm in clinical text. That compares to confirmation rates of 36%–100% reported in previous research regarding ICD-9-CM diagnoses.[6] Grades in this category showed high confidence and good agreement among our raters (Figure 1), with most events clearly judged to have documentation of self-harm intent and a small minority clearly judged to not have evidence of self-harm.

Among injuries and poisonings originally coded as accidental, only 8% had documentation of self-harm intent. Grades in this category showed some uncertainty, with more summary scores falling close to zero. We are not aware of other published data regarding evidence for self-harm among injuries and poisonings coded as accidental. We should emphasize that this 8% rate of documented self-harm intent applies only to the small proportion of all accidental injury and poisoning diagnosis codes selected for higher probability of self-harm (Table 1).

The 30% rate of documented self-harm among undetermined intent events in this sample differs from that in our prior evaluation of ICD-9-CM undetermined intent diagnoses in people with mental health diagnoses, where approximately 80% of events had documentation of self-harm intent.[14] This discrepancy may be explained by the decrease in undetermined intent diagnoses and increase self-harm diagnoses observed with the transition from ICD-9-CM to ICD-10-CM.[16] If the ICD-10-CM scheme led to more accurate coding of intent, then fewer self-harm events would remain in the undetermined category. Any naïve comparison of self-harm rates across

the transition from ICD-9-CM to ICD-10-CM would be seriously flawed. Events originally coded as having undetermined intent stand out for greater disagreement and lower confidence ratings (Figure 2). Greater uncertainty regarding these is expected, given that treating clinicians were not able to determine intent at the point of care.

Graders' evaluation of injuries without coding of intent showed high levels of agreement and confidence (Figure 1). Here again, we should emphasize that the 11% rate of documented self-harm intent seen in this sample applies only to injury codes specifically selected for higher expected probability of self-harm, and not to the 99% of injury codes not selected for review.

Events not originally coded as self-harm were sampled at different rates from different coding groups, over-sampling events in less common categories. Consequently, accuracy metrics such as recall or F1 score cannot be calculated directly from counts in Table 3. Those metrics can be estimated by applying the proportions of events with documentation of self-harm in this sample to total numbers of events in each coding category in a specific population. Supplementary Appendix B illustrates those calculations for the population of patients included in the pragmatic trial from which the majority of chart review events were sampled as well as a broader sample of health system members making outpatient mental health visits. Between those 2 samples, recall for self-harm codes ranged from 84.8% to 86.3%, and F_1 score ranged from 0.862 to 0.870. In both of those samples, injuries without coding of intent were estimated to contribute the largest number of missed self-harm events. Recall rate and F1 score would be lower in populations with larger proportions of undetermined intent, accidental intent, or uncoded intent events relative to the proportion originally coded as self-harm.
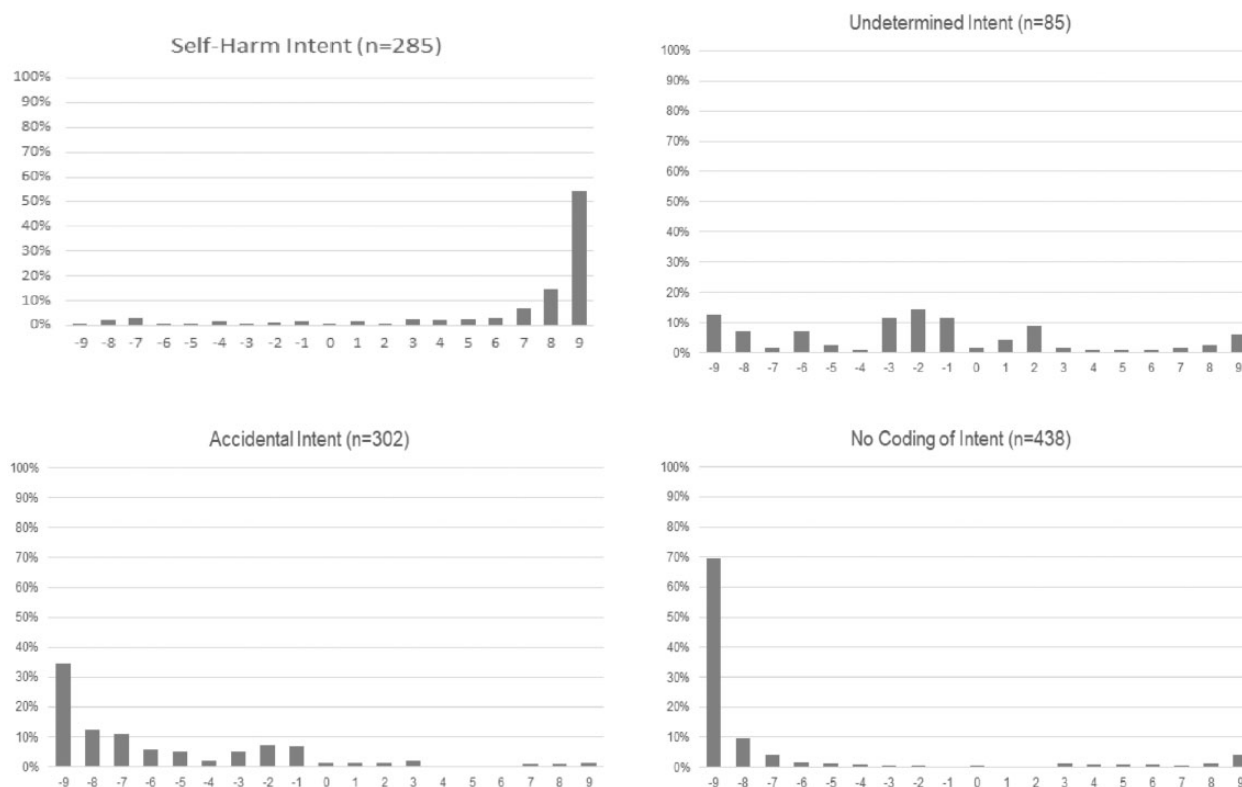
Figure 2. Distributions of weighted summary scores for documentation of self-harm intent according to original ICD-10-CM coding of intent. Scores range from −9 (all 3 graders classified as not self-harm with high confidence) to +9 (all 3 graders classified as self-harm with high confidence).
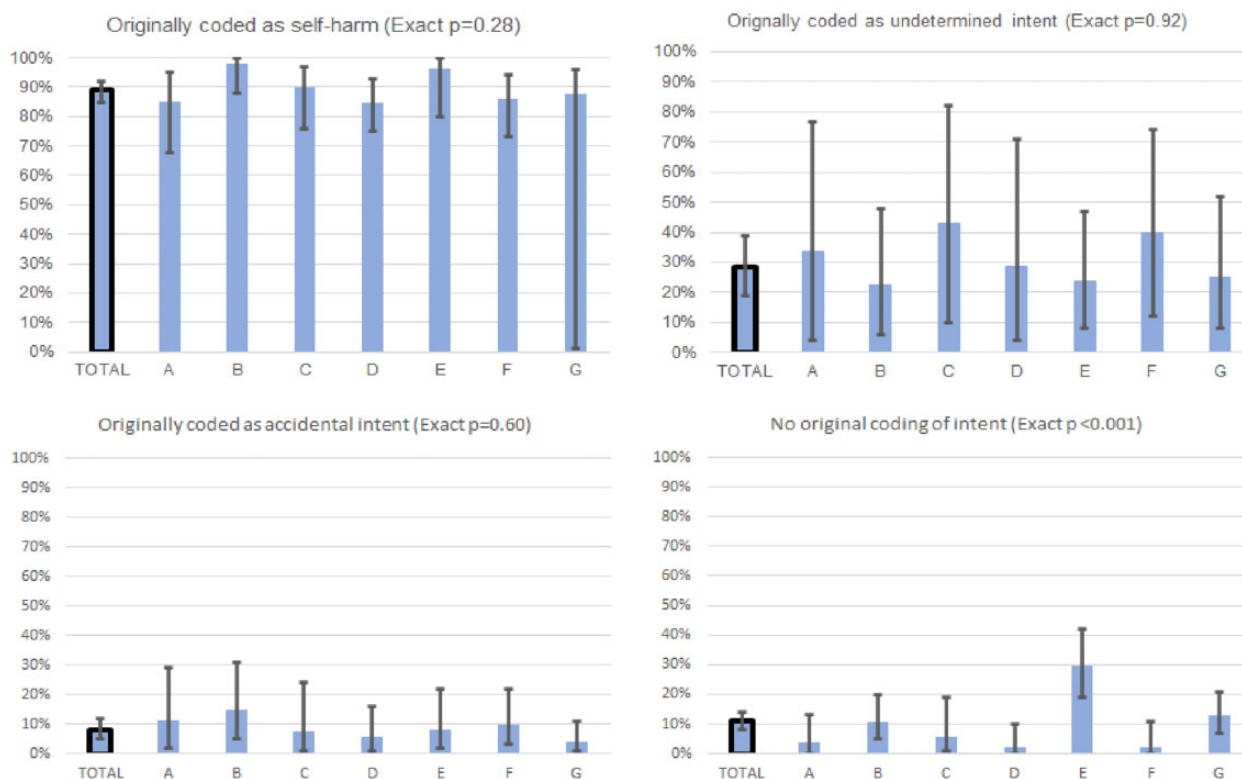
Figure 3. Proportion of events in each coding category with documentation of self-harm (based on summary score > 0), stratified by health system.

Using these findings to estimate the number of self-harm events not captured by ICD-10-CM intent coding would fail to count events among codes not selected for review (right column of Table 1). While some excluded codes (eg, spider bites) do not appear plausible as mechanisms of self-harm, some (eg, back or neck sprains, poisonings by unspecified drugs) could be consequences of self-harm. It was not feasible to review the large number of events in these categories, so we must acknowledge that some self-harm events would be excluded by our selection of specific codes and we cannot accurately estimate the number of true self-harm events not detected by our methods.

Patients in this sample had an expected risk of self-harm of approximately 4% over 18 months,[21] and these findings may not generalize to those with lower or higher risk. Among patients with lower risk of self-harm, we would expect both a lower confirmation rate for self-harm diagnoses and lower rates of missed events among injuries and poisonings not coded as self-harm. Among patients selected for low risk, diagnoses of self-harm may more often reflect coding errors than actual self-harm.[25] Conversely, we might expect a higher confirmation rate and higher rates of uncoded events among people at higher risk, such as those with a recent self-harm event or suicide attempt. Systematic assessment of misclassification among people at lower or higher risk would be needed to confirm those expectations.

Results were generally consistent across the 7 health systems, but we observed statistically significant variation (from 2% to 30%) in the proportion of injuries without coding of intent judged to have documentation of self-harm intent (Figure 1). Text extraction and rating procedures were identical across health systems, and raters were blinded to both health system and original coding. Consequently, this variation likely reflects true differences in coding. Injuries without coding of intent may include different proportions of true self-harm events in different health systems or care settings. The proportion of injury encounters without coding of intent does vary from state to state, and this could reflect geographic differences in failure to code intent when self-harm is suspected.[26,27]

These methods do not distinguish between self-harm with and without intent or expectation of death. We do not believe that text of clinical notes support making that distinction, and we believe that treating clinicians often find that distinction uncertain. Furthermore, patients receiving treatment for self-harm may be reluctant to report lethal intent in emergency department or inpatient settings if doing so leads to restrictive or coercive interventions.[28]

We should emphasize that this work does not consider other mechanisms by which encounter diagnosis codes may fail to identify self-harm events. First, people experiencing self-harm might not seek medical care, so no encounter would appear in health system records. Second, treating providers might not record any injury or poisoning diagnosis. Third, as noted above, some self-harm events might receive diagnoses (such as "other injury of unspecified body region") excluded from our record review process.

We should caution that our findings may not generalize to other health systems or care settings. While the ICD-10-CM taxonomy applies across the United States, recording of diagnoses is certainly subject to local influences.[9] Clinicians' coding of self-harm intent may be influenced by access to prior records and by how coding options are presented in the EHR. Diagnoses recorded by coding consultants may be influenced by local policies, practices, and EHR environments. Approximately 10% of encounters with reviewable diagnoses were not recorded in health system EHRs, and these findings may not generalize to that unreviewed subgroup.

CONCLUSIONS

Allowing for uncertainty regarding generalizability, we suggest the following implications of these findings. Data regarding coding of self-harm or undetermined intent under the ICD-9-CM and ICD-10-CM diagnostic systems should not be combined or treated as compatible. After the transition to ICD-10-CM, nearly all injuries and poisonings coded as self-harm have documentation of self-harm intent in clinical text. Researchers aiming to use health system data for population-based research should not generally include ICD-10-CM diagnoses of undetermined intent in definitions of self-harm. While 30% of events coded as having undetermined intent have clinical text indicating self-harm, relatively few events were coded as having undetermined intent. Consequently, this group would contribute minimally to the true total of self-harm events. The proportion of all true self-harm events that are mistakenly coded as accidental or that lack coding of intent is small. Those aiming to measure or monitor rates of self-harm in a specific population should expect that reliance on ICD-10-CM coding of self-harm will modestly under-estimate true numbers of events.

AUTHOR CONTRIBUTIONS

Conception and design: GES and SMS. Data acquisition: GES, JMB, RCR, GNC, JER, AB, and BB. Data analysis: GES, SMS, and CCS. Data interpretation: all authors. Drafting of manuscript: GES. Critical revision of manuscript: all authors.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

CONFLICT OF INTEREST STATEMENT

None declared.

DISCLAIMER

This article reflects the views of the authors and should not be construed to represent FDA's views or policies.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

1. Yard E, Radhakrishnan L, Ballesteros MF, *et al.* Emergency department visits for suspected suicide attempts among persons aged 12-25 years before and during the COVID-19 pandemic—United States, January 2019-May 2021. *MMWR Morb Mortal Wkly Rep* 2021; 70 (24): 888–94.

2. Hogan MF, Grumet JG. Suicide prevention: an emerging priority for health care. *Health Aff (Millwood)* 2016; 35 (6): 1084–90.

3. Simon GE, Shortreed SM, Rossom RC, *et al*. Effect of offering care management or online dialectical behavior therapy skills training vs usual care on self-harm among adult outpatients with suicidal ideation: a randomized clinical trial. *JAMA* 2022; 327 (7): 630–8.

4. Rossom RC, Simon GE, Beck A, *et al*. Facilitating action for suicide prevention by learning health care systems. *Psychiatr Serv* 2016; 67 (8): 830–2.

5. Sansing-Foster V, Haug N, Mosholder A, *et al*. Risk of psychiatric adverse events among montelukast users. *J Allergy Clin Immunol Pract* 2021; 9 (1): 385–93.e12.

6. Walkup JT, Townsend L, Crystal S, Olfson M. A systematic review of validated methods for identifying suicide or suicidal ideation using administrative or claims data. *Pharmacoepidemiol Drug Saf* 2012; 21: 174–82.

7. Swain RS, Taylor LG, Braver ER, Liu W, Pinheiro SP, Mosholder AD. A systematic review of validated suicide outcome classification in observational studies. *Int J Epidemiol* 2019; 48 (5): 1636–49.

8. National Center for Health Statistics. *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9_CM)*. In: Services DoHaH, ed. *Department of Health and Human Services*. Atlanta: Centers for Disease Control; 2021.

9. Lu CY, Stewart C, Ahmed AT, *et al*. How complete are E-codes in commercial plan claims databases? *Pharmacoepidemiol Drug Saf* 2014; 23 (2): 218–20.

10. Iribarren C, Sidney S, Jacobs DR Jr, Weisner C. Hospitalization for suicide attempt and completed suicide: epidemiological features in a managed care population. *Soc Psychiatry Psychiatr Epidemiol* 2000; 35 (7): 288–96.

11. Rhodes AE, Links PS, Streiner DL, Dawe I, Cass D, Janes S. Do hospital E-codes consistently capture suicidal behaviour? *Chronic Dis Can* 2002; 23 (4): 139–45.

12. Simon GE, Savarino J. Suicide attempts among patients starting depression treatment with medications or psychotherapy. *Am J Psychiatry* 2007; 164 (7): 1029–34.

13. Kumar P, Nestsiarovich A, Nelson SJ, Kerner B, Perkins DJ, Lambert CG. Imputation and characterization of uncoded self-harm in major mental illness using machine learning. *J Am Med Inform Assoc* 2020; 27 (1): 136–46.

14. Simon GE, Johnson E, Lawrence JM, *et al*. Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. *AJP* 2018; 175 (10): 951–60.

15. National Center for Health Statistics. *International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)*. In: Services DoHaH, ed. *Department of Health and Human Services*. Atlanta: Centers for Disease Control; 2021.

16. Stewart C, Crawford PM, Simon GE. Changes in coding of suicide attempts or self-harm with transition from ICD-9 to ICD-10. *Psychiatr Serv* 2017; 68 (3): 215.

17. Coleman KJ, Stewart C, Waitzfelder BE, *et al*. Racial-ethnic differences in psychiatric diagnoses and treatment across 11 health care systems in the mental health research network. *Psychiatr Serv* 2016; 67 (7): 749–57.

18. Ross TR, Ng D, Brown JS, *et al*. The HMO research network virtual data warehouse: a public data model to support collaboration. *eGEMs* 2014; 2 (1): 2.

19. Simon GE, Beck A, Rossom R, *et al*. Population-based outreach versus care as usual to prevent suicide attempt: study protocol for a randomized controlled trial. *Trials* 2016; 17 (1): 452.

20. Kroenke K, Spitzer RL, Williams JB, Lowe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. *Gen Hosp Psychiatry* 2010; 32 (4): 345–59.

21. Simon GE, Coleman KJ, Rossom RC, *et al*. Risk of suicide attempt and suicide death following completion of the Patient Health Questionnaire depression module in community practice. *J Clin Psychiatry* 2016; 77 (2): 221–7.

22. Clopper C, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934; 26 (4): 404–13.

23. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971; 76 (5): 378–82.

24. Fisher RA. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd; 1954.

25. Ludman EJ, Simon GE, Whiteside U, Richards JE, Pabiniak C. Reevaluating sensitivity of self-reported suicidal ideation. *J Clin Psychiatry* 2018; 79 (3): 17l12017.

26. Healthcare Cost and Utilization Project. *Injuries and External Causes: Reporting of Causes on the HCUP Inpatient Databases (SID), 2016-2019*. Bethesda, MD: Agency for Health Care Research and Quality; 2021.

27. Healthcare Cost and Utilization Project. *Injuries and External Causes: Reporting of Causes on the HCUP State Emergency Department Databases (SEDD), 2016-2019*. Bethesda, MD: Agency for Healthcare Research and Quality; 2021.

28. Richards JE, Whiteside U, Ludman EJ, *et al*. Understanding why patients may not report suicidal ideation at a health care visit prior to a suicide attempt: a qualitative study. *Psychiatr Serv* 2019; 70 (1): 40–5.