



Published in final edited form as:

Int J Med Inform. 2019 June ; 126: 19–25. doi:10.1016/j.ijmedinf.2019.03.002.

Using SNOMED CT-Encoded Problems to Improve ICD-10-CM Coding – a Randomized Controlled Experiment

Kin Wah Fung, MD, MS, MA¹, Julia Xu, MD, PhD¹, S. Trent Rosenbloom, MD, MPH², and James R. Campbell, MD³

¹National Library of Medicine, Bethesda, MD

²Vanderbilt University, Nashville, TN

³University of Nebraska Medical Center, Omaha, NE

Abstract

Objective: Clinical problems in the Electronic Health Record that are encoded in SNOMED CT can be translated into ICD-10-CM codes through the NLM’s SNOMED CT to ICD-10-CM map (NLM Map). This study evaluates the potential benefits of using the map-generated codes to assist manual ICD-10-CM coding.

Methods: De-identified clinic notes taken by the physician during an outpatient encounter were made available on a secure web server and randomly assigned for coding by professional coders with usual coding or map-assisted coding. Map-assisted coding made use of the problem list maintained by the physician and the NLM Map to suggest candidate ICD-10-CM codes to the coder. A gold standard set of codes for each note was established by the coders using a Delphi consensus process. Outcomes included coding time, coding reliability as measured by the Jaccard coefficients between codes from two coders with the same method of coding, and coding accuracy as measured by recall, precision and F-score according to the gold standard.

Results: With map-assisted coding, the average coding time per note reduced by 1.5 minutes ($p=0.006$). There was a small increase in coding reliability and accuracy (not statistical

Correspondence and reprints: Kin Wah Fung, Building 38A, Rm9S918, MSC-3826, National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894, Telephone: 301 – 827 5001, Fax: 301 – 496 0663, kwfung@nlm.nih.gov.

Contributorship Statement

KWF and JRC conceived and designed the study. JRC prepared the clinical notes and recruited the coders. KWF coordinated the coding experiment, data collection and consensus building to establish the gold standard. JX and TR performed the failure analysis. KWF carried out the data analysis. KWF drafted the manuscript and all authors contributed substantially to its revision.

Author statement

I, Kin Wah Fung, as corresponding author, am submitting the manuscript titled “Using SNOMED CT-Encoded Problems to Improve ICD-10-CM Coding – a Randomized Controlled Experiment” for consideration of publication in the International Journal of Medical Informatics I confirm that this submission is a new, original work that has not been previously submitted, published in whole or in part, or simultaneously submitted for publication in another journal.

Competing Interests Statement

The authors have no competing interests.

Conflicts of Interest Statement

The authors have no conflicts of interest.

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

significant). The benefits were more pronounced in the more experienced than less experienced coders. Detailed analysis of cases in which the correct ICD-10-CM codes were not found by the NLM Map showed that most failures were related to omission in the problem list and suboptimal mapping of the problem list terms to SNOMED CT. Only 12% of the failures was caused by errors in the NLM Map.

Conclusion: Map-assisted coding reduces coding time and can potentially improve coding reliability and accuracy, especially for more experienced coders. More effort is needed to improve the accuracy of the map-suggested ICD-10-CM codes.

Keywords

SNOMED CT; ICD-10-CM; inter-terminology mapping; administrative codes; coding quality

INTRODUCTION

We are entering the era of interoperable electronic health records (EHR) and witnessing the genesis of the Learning Healthcare System envisioned by the U.S. National Academy of Medicine (previously known as Institute of Medicine).¹ Guided by the Meaningful Use of EHR incentive program of the U.S. Centers for Medicare and Medicaid Services (CMS) and its successor Promoting Interoperability,^{2,3} an architecture for data management within the EHR has been delineated.⁴ We are seeing increasing quantity of coded electronic data available to guide clinical care and harvested to serve public health and research. For big data analytics, administrative codes generated for insurance claims are one of the most frequently used data sources because of their volume and ubiquity. In the U.S., ICD-10-CM has been the primary administrative code set used for transaction activities such as insurance reimbursement since 2015. However, the quality of ICD codes has been called into question by various studies over the years. While individual study results may vary, a recurring message is that there is room for improvement. For example, a study by the National Academy of Medicine on the reliability of hospital discharge coding showed that only 65% agreed with independent re-coding.⁵ Another study by Hsia et al found a coding error rate of 20%.⁶ Among other studies, the typical error rate is about 25 – 30%, with low agreement between coders.^{7–11}

In the U.S., there has been substantial growth in the use of EHRs in the past decade due to incentive programs and other factors.^{12,13} There is an accompanying increase in the amount of clinical information captured in a structured format and encoded in clinical terminology standards such as SNOMED CT, LOINC and RxNorm. There are important distinctions between clinical terminologies and administrative classifications (such as ICD-10-CM) and each should be used according to their designated purpose. Compared to ICD-10-CM, SNOMED CT is better suited to support clinical care because of better content coverage, clinical-orientation, flexible data entry and retrieval capabilities leveraging the underlying logical construct.^{14–23} According to the EHR certification criteria, problem list entries in the EHR are required to be encoded in SNOMED CT.²⁴ SNOMED CT is steadily gaining momentum as the emerging international clinical terminology standard.^{25,26} The number of member countries has more than tripled (from 9 to 35) since the establishment of the

International Health Terminology Standards Development Organisation (IHTSDO, now called SNOMED International) in 2007.²⁷

The availability of encoded clinical problems in the EHR offers new opportunities in ICD-10-CM coding. The coarse-grained ICD-10-CM codes can, in most cases, be inferred from the fine-grained SNOMED CT codes, together with consideration of co-morbidity, patient demographics and other variables. This will not only save coding time, but can potentially improve ICD-10-CM coding quality. To facilitate the translation between SNOMED CT and ICD-10-CM codes, the U.S. National Library of Medicine (NLM) publishes a map from SNOMED CT to ICD-10-CM (hereafter called the NLM Map). The NLM Map, first released in 2012, is regularly updated and expanded in coverage.²⁸ The September 2018 version of the map covers 123,260 SNOMED CT concepts. The NLM Map is a rule-based map to cater for the coding rules of ICD-10-CM, which may point to different codes depending on the patient's age, gender and other factors. For example, the SNOMED CT concept *Sleep apnea (73430006)* can map to either *P28.3 Primary sleep apnea of newborn* or *G47.30 Sleep apnea, unspecified* depending on whether the patient is a neonate or an adult. One intended use of the map is to support real-time, interactive generation of ICD-10-CM codes for use by clinicians, as exemplified by the I-MAGIC tool.²⁹ Another use of the map is to provide assistance to the professional coder. A typical scenario can be like this: Patient Mr. Jones is being seen by Dr. Jane in the clinic. Dr. Jane has maintained a historical diagnosis and health-related problem list coded in SNOMED CT while caring for Mr. Jones and updates the problem list, selecting the problems which are addressed during today's visit. For example, Mr. Jones notices recent thinning of his hair and Dr. Jane adds 'hair loss' to the problem list for this visit. In the backend database, the term 'hair loss' is mapped to the SNOMED CT concept *Alopecia (56317004)*. When the clinical note is passed to the professional coder in the billing office for coding, the EHR uses the NLM Map to look up ICD-10-CM codes corresponding to the constellation of SNOMED CT problems selected by Dr. Jane. The coding software displays the ICD-10-CM code *Nonscarring hair loss, unspecified (L65.9)* to the coder as suggestion for the newly added problem. The coder decides whether to accept the suggestions based on their professional judgement. We hypothesize that such an approach can save coding time, improve coding consistency and accuracy. We report our findings in a randomized controlled experiment to study the potential benefits of map-assisted coding. As far as we know, this is the first formal study of map-assisted ICD-10-CM coding.

METHODS

Preparation of clinical notes

A research protocol was developed by the investigators, submitted and approved by the Institutional Review Boards of the University of Nebraska Medical Center and the National Institutes of Health. University of Nebraska Medical Center is a state institution with all enterprise profits used to operate the facility. University of Nebraska Medical Center employs the Epic© EHR system at all sites. Epic© supports an integrated data warehouse Clarity© which daily extracts a copy of the EHR data for use in quality management and analytics. EHR data gathered from Clarity© for this study included full text notes recorded

by the physician, problem list data personally entered by the physician, and the subsequent billing codes and diagnoses developed by the billing office on their review of the record. One investigator (JRC) obtained a convenience sample of 102 full text clinical encounter notes of moderate to high complexity from the enterprise data warehouse at Nebraska, all representing ambulatory care encounters for adult patients having been documented and billed at level 4 subsequent care (denoted by the CPT code *99214 Established patient office or other outpatient, visit typically 25 minutes*) during the previous calendar year. We chose this level of service code to identify a cohort of patients who were seeking continuity care within the practice and had an encounter which required detailed history, detailed examination and medical decision making of moderate complexity. These notes were selected from the internal medicine outpatient clinic and represented care events across a sample of providers working in the clinic. The notes included chief complaints, medical histories, physical examinations, medication lists, laboratory results as recorded by the clinician, diagnostic assessments and plans. In addition, all problem list terms picked by the physician during encounter sign out, with their associated SNOMED CT encodings provided by the EHR vendor, were retrieved and included in the study data. The problem terms were selected from a problem list vocabulary maintained by the EHR vendor, and their SNOMED CT encodings were invisible to the clinician. Based on the SNOMED CT codes and other patient parameters such as age and gender, we used the NLM Map to suggest candidate ICD-10-CM codes to the professional coders participating in this study. Data quality within the EHR was not studied in this project and is not a subject of this report which is developed as a retrospective analysis of the encounter events and the value proffered by the SNOMED CT maps.

In compliance with the research protocol approved by the Institutional Review Boards, one author (JRC) manually reviewed and anonymized the text notes, removing all patient identifying information and institutional identifiers, while preserving gender and rounding ages to 10 years. Another author (KWF) validated the anonymization to ensure no patient identifying information remained. The notes were hosted on a secure web server at NLM for the period of the study with user account and password control.

Usual vs. map-assisted coding

Four professional coders with proper certification for ICD-10-CM coding by the American Health Information Management Association (AHIMA) or equivalent organizations were recruited from the coding office staff at Nebraska. The 102 notes were divided randomly into 17 blocks of 6 notes each. Using the Latin square method, each note in a block was assigned to two coders for coding by the usual method and the other two coders by the map-assisted method. The 6 notes in a block covered all possible combinations of coder and method of coding. The coders reviewed the de-identified notes via a secure website, as shown in figure 1. For notes assigned for usual coding, only the problem terms were shown in the box on the right. For map-assisted coding, the coders would also see the ICD-10-CM code suggestions based on the NLM Map. Most problem terms had one suggested ICD-10-CM code, while a small number of terms had zero or multiple ICD-10-CM code suggestions.

The coders were instructed to review the clinical notes and encode all clinical problems relevant to the encounter, not only those specific for reimbursement. Coders could use the problem terms and the ICD-10-CM code suggestions as additional guidance, but were not obliged to follow them. They could use any electronic or paper-based reference material that they would normally use. To avoid typos, the coders picked from a picklist generated by auto-completion as they typed in the code. For this study, no principal diagnosis was assigned and code sequence was not important. The time from displaying a note to clicking the “Done” button was captured as the coding time. Once a note was finished, the coder could not return to it. The coders were instructed to complete the coding of a note without interruption to ensure the accuracy of coding time measurement. They could use the “Pause” button if they had to pause for some reason. We also recorded any accidental interruption of coding (e.g., browser stopped working, network connection severed), which would lead to erroneous time keeping, as the timer was reset when a note was displayed. To minimize the learning bias as the coders were getting used to the tool, the notes were presented to all coders in exactly the same order.

Gold standard by consensus with Delphi method

A gold standard list of codes was established for each note by the coders themselves using the Delphi method after the coding experiment was completed. For each note, codes entered by three or more coders were accepted as final. Undecided codes were defined as either a code entered by exactly two coders, or two or more related codes entered by different coders. For example, if one coder entered *Chronic lymphocytic leukemia of B-cell type not having achieved remission (C91.10)*, while another coder entered *Chronic lymphocytic leukemia of B-cell type in remission (C91.11)*, both *C91.10* and *C91.11* were considered as related codes and kept for voting. Related codes must share at least the first three digits, and were confirmed to be related by manual review. Codes entered by only one coder with no related codes from other coders were discarded. We chose to discard these codes because including all codes in the Delphi process would considerably increase the number of codes that needed to be voted on, the rounds of voting and the codes that would remain in a stalemate at the end of voting. After all, it was unlikely that a code picked by only one coder and not any other coders would end up being correct. Undecided codes were put in a spreadsheet for anonymous voting by the coders. Undecided codes receiving three or more votes were accepted to the gold standard. Codes with exactly two votes remained undecided and those with less than two votes were discarded. Undecided codes were put up for a further rounds of voting until a decision was made. If a stalemate occurred after several rounds of voting, a fifth coding specialist not involved in the coding experiment would cast the deciding vote.

Data analysis

Coding time, consistency and accuracy—We analyzed the coding time for each note by two-way analysis of variance (ANOVA), using coder and method of coding (usual or map-assisted) as independent variables. For coding consistency (or reliability), we calculated a similarity coefficient (Jaccard) for each clinical note between the lists of codes from two coders using the same method of coding. The Jaccard coefficient was calculated as the number of concurrent codes divided by the total number of unique codes in the two lists. We

chose the Jaccard coefficient over other interrater agreement measures such as Cohen's kappa or Krippendorff's alpha because the number of ICD-10-CM codes assigned to each note was variable, which would be problematic in the calculation and interpretation of Cohen's kappa or Krippendorff's alpha. We compared the Jaccard coefficients for the two coding methods by the non-parametric Mann-Whitney U Test. For coding accuracy, we measured the recall, precision and F-score for each coder's codes for a clinical note compared to the gold standard. Recall was the percentage of gold standard codes found by a coder. Precision was the percentage of a coder's codes that were in the gold standard. F-score was the harmonic mean of recall and precision. We analyzed the distribution of the recall, precision and F-score by two-way ANOVA using coder and method of coding as independent variables. To see the effect of coder experience – the number of years of coding – on the outcomes, we did a secondary analysis of coding time and accuracy according to coder experience. All statistical analysis was done by the IBM® SPSS® Statistics package version 21.

Failure analysis—For map-assisted coding, there are three steps to arrive at the correct candidate ICD-10-CM code. (Figure 2) First, the physician needs to select the relevant problem term. Second, the problem term is mapped to the appropriate SNOMED CT code. Third, the NLM Map finds the correct ICD-10-CM code based on the SNOMED CT code. We define failure as a gold standard ICD-10-CM code not present among the map-suggested candidate codes. All failure cases were reviewed by two authors (JX and STR) independently to assign a reason for failure, which could be due to any of the three steps. The results from the two reviewers were collated and discrepancies discussed to arrive at a consensus. A third reviewer (KWF) cast a deciding vote if consensus could not be reached.

To focus on the impact of the NLM Map - excluding upstream reasons of failure (those not caused by the NLM Map, i.e., problems 1 to 3 in figure 2) - we adjusted the gold standard by excluding codes that could not be reached because of upstream problems. We assessed the accuracy of the map-suggested codes based on the original and adjusted gold standards.

RESULTS

Creation of the gold standard by consensus

Of the 102 clinical notes, one note was displayed incorrectly due to a software glitch and was excluded from analysis. On average, each note had 5.6 (SD 2) problem terms, which were mapped to 6.6 (SD 2.2) SNOMED CT codes, resulting in 6.2 (SD 2.2) map-suggested ICD-10-CM codes. The coders entered 9.3 (SD 3.3) codes per note on average. The Delphi consensus process is summarized in table 1. In the final gold standard, there were 5.5 (SD 2) codes per note.

Coding time, consistency and accuracy

The results are summarized in table 2. Out of 404 coding times measured, 3 were excluded because the same note was viewed in more than one browser session, indicating accidental interruption of coding. Overall, there was a reduction of 1.5 minute of coding time per note for map-assisted compared to usual coding, which was statistically significant (two-way

ANOVA $p=0.006$). However, this benefit was not observed across all coders. The biggest reduction in coding time was observed in coder 4. The difference of coding time between coders did not reach statistical significance (two-way ANOVA $p=0.056$).

For coding consistency, as measured by the Jaccard similarity coefficients between codes from different coders with the same method of coding, map-assisted coding resulted in a slightly higher consistency (average increase of 0.07) compared to usual coding. However, the difference was not statistically significant (Mann-Whitney U test $p=0.109$).

Regarding coding accuracy, three coders (coders 1, 3 and 4) had small improvements in recall, precision and F-score with map-assisted coding while coder 2 suffered a slight drop. The overall mean recall, precision and F-score for all coders showed a small increase with map-assisted coding but the difference was not statistically significant.

To examine the effect of coding experience, we separated the coders into two groups - more experienced (over 15 years of coding, coders 1 and 4) and less experienced (less than 5 years, coders 2 and 3). (Table 3) The results seem to suggest that map-assisted coding may benefit experienced coder more by shortening coding time and improving accuracy.

Failure analysis

Issues with the problem terms entered by the physician accounted for the majority of the failure cases. (Table 4) In 39% of cases, the relevant problem term was missing. In 21% of cases, a suboptimal problem term was entered which could not lead to the correct ICD-10-CM code. Omission of laterality was a common problem and accounted a third of cases in this category. For example, a patient with right hip osteoarthritis was coded as *Osteoarthritis of hip*. Had the physician used *Osteoarthritis of the right hip*, the correct ICD-10-CM code, *Unilateral primary osteoarthritis, right hip (M16.11)*, would have been found by the NLM Map. In 28% of cases, the vendor's SNOMED CT map of the problem term was suboptimal. For example, the problem term *Nevus* was mapped to the SNOMED CT concept *Skin lesion (95324001)*. Had it been mapped to *Non-neoplastic nevus (195381005)*, the correct ICD-10-CM code *Nevus, non-neoplastic (I78.1)* would have been found by the NLM Map. In the remaining 12% of cases, the NLM Map was the reason of failure. For example, the SNOMED CT concept *Needs influenza immunization (185903001)* was mapped to *Underimmunization status (Z28.3)* in the NLM Map, but *Encounter for immunization (Z23)* was the correct code.

Accuracy of the codes suggested by the NLM Map

Based on the failure analysis, we adjusted the gold standard by excluding 143 codes which could not be reached by the NLM Map due to upstream problems. We estimated the accuracy of the map-suggested codes (as if they were entered by an "NLM auto-coder"), using both the original and adjusted gold standards. (Table 5) When judged by the original gold standard, the raw performance of the auto-coder was inferior to the average human coder. When judged by the adjusted gold standard, i.e., focusing on cases where there was a relevant problem term which was mapped correctly to SNOMED CT, the auto-coder would be comparable to a human coder, as shown by the similar F-scores (0.81 vs. 0.79).

DISCUSSION

With increasing use of the EHR in both hospitals and small practices related to the Meaningful Use incentive program, more and more patient data encoded in clinical terminologies are becoming available. One of the goals of the NLM's SNOMED CT to ICD-10-CM map is to enable the re-use of SNOMED CT encoded clinical data for the generation of ICD-10-CM codes. We hypothesize that using encoded clinical data to suggest administrative codes can lead to faster and better coding. Our results are somewhat mixed but serve to highlight some interesting points.

There is indeed considerable variability in coding as shown by the meagre similarity scores among coders. On average, only about half of the codes from two coders agree (average Jaccard coefficient 0.53 for usual coding). This level of agreement is also borne out by the observation during the Delphi process to establish the gold standard. Before the first voting, only 433 (46%) out of 942 codes were accepted into the gold standard because they came from two or more coders. This confirms the need to improve the reliability of manual coding. In our study, we showed that there is potential for map-assisted coding to improve the concordance between coders, but to reach statistical significance will probably need a bigger sample and improvement in the quality of the map-suggested codes.

Overall, map-assisted coding is likely to result in faster coding – we found an average reduction of 1.5 minutes (16%) in the coding time of a clinical note of moderate to high complexity. However, this effect is not observed uniformly across all coders. It seems that more experienced coders are more likely to benefit. This is understandable since they are more proficient in distinguishing good and bad suggestions, whereas less experienced coders need to spend extra time looking up and considering the suggested codes which will slow them down. Also important is that for more experienced coders, at least the shorter coding time does not negatively impact the accuracy of coding (the recall and precision of map-assisted coding was actually higher but the difference did not reach statistical significance). On the other hand, one caveat is that bad code suggestions can potentially confuse less experienced coders, leading to reduction in accuracy.

For map-assisted coding to confer real advantage, the map-suggested codes need to be of high enough quality. Otherwise, irrelevant or incorrect map-suggested codes can slow down and mislead coders. In our study, the map-suggested codes have meagre recall and precision of 0.71 and 0.63 respectively. In order to arrive at the correct code suggestion, the sequential steps beginning with the clinician picking the problem terms, to mapping from the problem term to SNOMED CT, then using the NLM Map to find the ICD-10-CM target, must all work properly. In reality, error can occur in any of these steps. Our failure analysis reveals that, upstream problems that occur before the use of the NLM Map account for the majority of errors. The NLM Map itself only accounts for 12% of the reviewed cases. Some of the upstream data issues can be mitigated at the point of physician data entry. Suboptimal mapping between the problem list terms and SNOMED CT can be reduced if the maps can be reviewed periodically, as more granular terms are added to newer versions of SNOMED CT. It is also possible to prompt the physician for additional information if the term they pick is not specific enough for ICD-10-CM coding. For example, in the I-MAGIC demo

tool, the users are prompted to enter additional information such as laterality, episode of care and trimester where appropriate.²⁹ In our study, missing laterality information accounts for 17% of failure cases. This is related to the fact that until recently, most SNOMED CT concepts did not specify laterality. There has been a change in the editorial policy of SNOMED CT and more and more lateralized concepts are being added. This will lead to a reduction in this type of error in future. There are some cases in which the NLM Map suggests a code that is not correct. For example, while *Underimmunization status (Z28.3)* is semantically a correct map for *Needs influenza immunization (185903001)*, the required code is *Encounter for immunization (Z23)*. We shall improve the NLM Map based on these findings.

Finally, there is the question of whether ICD-10-CM coding can be fully automated based on the clinical information in the EHR. In our study, the raw accuracy of map-generated coding is inferior to the average coder, which means that the map-suggested codes can only be used to assist, but not replace manual coding. Theoretically, if all upstream problems could be fixed, the map-suggested codes would be as accurate as the average coder, and automatic coding could be considered an alternative. But until then, map-suggested codes should be reviewed manually before being used for reimbursement and other purposes.

We recognize the following limitations in our study. The convenience sample of clinical notes came from one internal medicine clinic and may not be representative of patients from other medical specialties or institutions. We did not consider the order of the ICD-10-CM codes and did not ask the coders to assign a principal diagnosis, which is normally done in reimbursement coding. Two authors (KWF and JX) are involved in the production of the NLM Map. JRC was involved in the initial conceptualization of the NLM Map. To avoid subjective bias in the failure analysis, another author (TR) who is not involved in the NLM mapping project did an independent review and discrepancies were discussed until consensus was reached.

CONCLUSION

We performed a randomized, controlled experiment in ICD-10-CM coding using de-identified patient notes to study the effects of map-assisted coding using the NLM's SNOMED CT to ICD-10-CM map. There is significant reduction in coding time and possibly slight improvements in coding reliability and accuracy. More experienced coders tend to benefit more from map-assisted coding than less experienced coders. Most of the cases in which the correct code is not found by the NLM Map are not due to the map itself, but attributable to the problem term or the translation of the problem term to SNOMED CT codes. After adjusting for these factors, the map-suggested codes could be comparable in accuracy to a human coder.

Acknowledgements

We would like to thank all the coders who participated in this study. Also, we thank Paul Lynch and Frank Tao for creating the web application to host the de-identified clinical notes.

Funding Statement

This research was supported in part by the Intramural Research Program of the NIH, National Library of Medicine.

Research at the University of Nebraska was supported by PCORI CDRN-1501-26643 CDRN 04: The Greater Plains Collaborative and the UNMC Research Information Technology Office.

References:

1. Institute of Medicine (IOM). 2007 The Learning Healthcare System: Workshop Summary. Washington, DC: The National Academies Press.
2. US Office of the National Coordinator for Health Information Technology, Department of Health and Human Services: Meaningful use website <http://www.healthit.gov/policy-researchers-implementers/meaningful-use>.
3. Promoting Interoperability, Centers for Medicare & Medicaid Services Available from: <https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/ehrincentiveprograms/>.
4. The Interoperability Standards Advisory 2017, Office of the National Coordinator for Health Information Technology (ONC) Available from: <https://www.healthit.gov/isa/>.
5. Institute of Medicine, Reliability of Hospital Discharge Records, Report of a study, National Academy of Sciences, Washington DC, 1977.
6. Hsia DC, et al., Accuracy of diagnostic coding for Medicare patients under the prospective-payment system. *N Engl J Med*, 1988 318(6): p. 352–5. [PubMed: 3123929]
7. Surjan G, Questions on validity of International Classification of Diseases-coded diagnoses. *Int J Med Inform*, 1999 54(2): p. 77–95. [PubMed: 10219948]
8. O'Malley KJ, et al., Measuring diagnoses: ICD code accuracy. *Health Serv Res*, 2005 40(5 Pt 2): p. 1620–39. [PubMed: 16178999]
9. Dixon J, et al., Assessment of the reproducibility of clinical coding in routinely collected hospital activity data: a study in two hospitals. *J Public Health Med*, 1998 20(1): p. 63–9. [PubMed: 9602451]
10. Hogan WR and Wagner MM, Accuracy of data in computer-based patient records. *J Am Med Inform Assoc*, 1997 4(5): p. 342–55. [PubMed: 9292840]
11. Stausberg J, et al., Reliability of diagnoses coding with ICD-10. *Int J Med Inform*, 2008 77(1): p. 50–7. [PubMed: 17185030]
12. Henry J PY, Searcy T et al., Adoption of Electronic Health Record Systems among US Non-Federal Acute Care Hospitals: 2008–2015 ONC Data Brief. Washington DC: Office of the National Coordinator for Health Information Technology 35: p. 1–9.
13. Adler-Milstein J, et al., Electronic Health Record Adoption In US Hospitals: Progress Continues, But Challenges Persist. *Health Aff (Millwood)*, 2015 34(12): p. 2174–80. [PubMed: 26561387]
14. Chute CG, et al., The content coverage of clinical classifications. For The Computer-Based Patient Record Institute's Work Group on Codes & Structures. *J Am Med Inform Assoc*, 1996 3(3): p. 224–33. [PubMed: 8723613]
15. Campbell JR, et al., Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity. CPRI Work Group on Codes and Structures. *J Am Med Inform Assoc*, 1997 4(3): p. 238–51. [PubMed: 9147343]
16. Yasini M, et al., Comparing the use of SNOMED CT and ICD10 for coding clinical conditions to implement laboratory guidelines. *Stud Health Technol Inform*, 2013 186: p. 200–4. [PubMed: 23542998]
17. Chen JW, Flaitz C, and Johnson T, Comparison of accuracy captured by different controlled languages in oral pathology diagnoses. *AMIA Annu Symp Proc*, 2005: p. 918. [PubMed: 16779205]
18. McClay JC and Campbell J, Improved coding of the primary reason for visit to the emergency department using SNOMED. *Proc AMIA Symp*, 2002: p. 499–503. [PubMed: 12463874]
19. Elkin PL, et al., A randomized controlled trial of the accuracy of clinical record retrieval using SNOMED-RT as compared with ICD9-CM. *Proc AMIA Symp*, 2001: p. 159–63. [PubMed: 11825173]

20. Vardy DA, Gill RP, and Israeli A, Coding medical information: classification versus nomenclature and implications to the Israeli medical system. *J Med Syst*, 1998 22(4): p. 203–10. [PubMed: 9690178]
21. Liu H, Waghlikar K, and Wu ST, Using SNOMED-CT to encode summary level data - a corpus analysis. *AMIA Summits Transl Sci Proc*, 2012 2012: p. 30–7. [PubMed: 22779045]
22. Fung KW, McDonald C, and Srinivasan S, The UMLS-CORE project: a study of the problem list terminologies used in large healthcare institutions. *J Am Med Inform Assoc*, 2010 17(6): p. 675–80. [PubMed: 20962130]
23. Fung KW and Xu J, An exploration of the properties of the CORE problem list subset and how it facilitates the implementation of SNOMED CT. *J Am Med Inform Assoc*, 2015 22(3): p. 649–58. [PubMed: 25725003]
24. Health Information Technology Certification Criteria, U.S. Health and Human Services Department on 10/16/2015. Available from: <https://www.federalregister.gov/articles/2015/10/16/2015-25597/2015-edition-health-information-technology-health-it-certification-criteria-2015-edition-base>.
25. Lee D, et al., A survey of SNOMED CT implementations. *J Biomed Inform*, 2013 46(1): p. 87–96. [PubMed: 23041717]
26. Lee D, et al., Literature review of SNOMED CT use. *J Am Med Inform Assoc*, 2014 21(e1): p. e11–9. [PubMed: 23828173]
27. SNOMED International Member Countries. Available from: <http://www.snomed.org/members/>.
28. SNOMED CT to ICD-10CM Map, http://www.nlm.nih.gov/research/umls/mapping_projects/snomedct_to_icd10cm.html.
29. I-MAGIC demo tool, NLM. Available from: <https://imagic.nlm.nih.gov/imagic/code/map>.

SUMMARY TABLE

What was already known on the topic:

- ICD-10-CM codes are often used in data mining and analytics because they are ubiquitous, but the quality of coding has been called into question
- With increased usage of electronic health records, more encoded clinical data become available

What this study added to our knowledge:

- It is feasible to use existing SNOMED CT encoded clinical problems to assist in ICD-10-CM coding through the use of the NLM SNOMED CT to ICD-10-CM Map
- Map-assisted coding can potentially lead to faster, more reliable and accurate coding, especially for more experienced coders
- To reap the benefits of map-assisted coding, the quality of the map-suggested codes needs to be improved

Highlights:

- ICD-10-CM coding quality has been called into question
- SNOMED CT encoded clinical problems in the EHR can be used to assist ICD-10-CM coding
- Map-assisted coding reduces coding time and can potentially improve coding reliability and accuracy for experience coders

MapStudy NLM Map-Assisted ICD-10-CM Coding Study - Task 27 of 102 Hello kinfung2 Logout

Assign ICD-10-CM codes to the following text:

Note #27

Subjective

The patient is a 60-year-old woman, who is here today for followup of hyperlipidemia, Meniere disease, hypertension, and anxiety.

Earlier this year in January, the patient came into the clinic having panic attack. She had multiple stressors at that time but the predominant one was the recent death of her mother. She was having problems with the grieving process. I did have her see [DOCTOR] at the [LOCATION] in Psychology. She met with her several times and has done very good with regard to her anxiety. I had prescribed her some alprazolam at that time and she just takes this on a rare basis. She does still admit to being sad at times when thinking about the death of her mother.

With regard to her hypertension, she actually takes hydrochlorothiazide and triamterene for this and also Meniere disease, which has been fairly stable. She did have some dyslipidemia and is not currently on any medication at this point.

With regard to health care maintenance, she does not need cervical cancer screening given her hysterectomy. She had a colonoscopy in 2010, which was difficult due to the tortuous nature of her colon. It was recommended she had a virtual colonoscopy, which was not covered by insurance and other possibility is barium enema. She has had no blood in her stools. She does agree to do stool guaiac cards. Last mammogram of note was in 2012. Vaccinations, she is up-to-date with regard to Tdap and has been given prescription for a shingles vaccine in the past, but has not done this yet.

PAST MEDICAL HISTORY:

1. Meniere's disease.
2. Hysterectomy with bilateral oophorectomy for fibroids.
3. Hypertension.

Please begin. Click **Pause** if needed.

Diagnostic Problems & Map-Suggested ICD-10-CM Codes:

Problem	Code	Description
Hyperlipidemia	E78.5	Hyperlipidemia, unspecified
Hyperlipidemia	R79.89	Other specified abnormal findings of blood chemistry
Hypertension	I10	Essential (primary) hypertension
Need for pneumococcal vaccination	Z28.3	Underimmunization status

Record your ICD-10-CM codes here:

1. I10 - Essential (primary) hypertension

[Done](#)

Figure 1. Screenshot of the secure online coding tool. The de-identified clinical note is on the left. The box on the right contains the problem terms entered by the physician and the map-suggested ICD-10-CM codes, which are not visible in the usual coding mode.

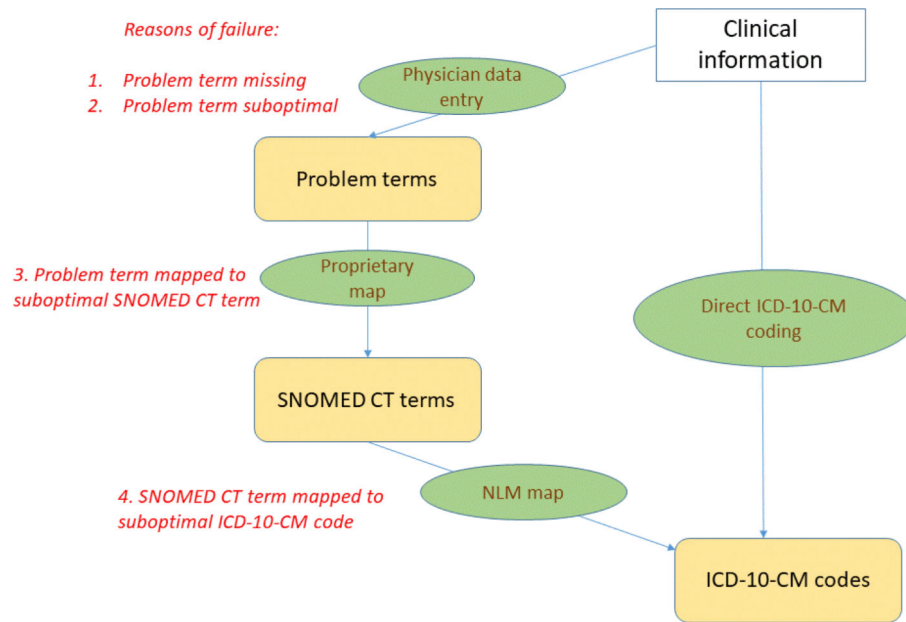


Figure 2. Comparison of the two methods of coding. The left side shows map-assisted coding and the right side usual coding. The correct code can be missed in the map-suggested codes because of problems related to problem term entry by the physician, the mapping of the problem term to SNOMED CT or the NLM SNOMED CT to ICD-10-CM map.

Table 1.

Distribution of ICD-10-CM codes in the Delphi consensus process. Undecided codes moved on to the next round of voting. After the second round of voting, the deciding vote was cast by a fifth coding specialist.

	Accepted to gold standard (> 2 coders)	Undecided (2 coders or has related code)	Rejected (only 1 coder, no related codes)
Before voting	433	256	253
Result of first voting	109	34	113
Result of second voting	8	9	17
Result of deciding vote	3	0	6
Final results	553	0	389

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Coding time, consistency and accuracy for usual and map-assisted coding

Metric	Usual coding	Map-assisted coding
	Mean (SD)	Mean (SD)
Coding time (minute)		
Coder 1	8.51 (5.86)	8.62 (6.89)
Coder 2	8.68 (3.63)	9.13 (4.95)
Coder 3	8.85 (4.85)	8.02 (4.6)
Coder 4	9.78 (8.11)	4.11 (2.61)
All coders	8.95 (5.81)	7.48 (5.34)
Coding consistency (Jaccard coefficient)		
All coders	0.53 (0.27)	0.6 (0.27)
Coding accuracy		
Recall		
Coder 1	0.83 (0.2)	0.85 (0.2)
Coder 2	0.82 (0.19)	0.8 (0.2)
Coder 3	0.84 (0.16)	0.86 (0.16)
Coder 4	0.72 (0.25)	0.76 (0.22)
All coders	0.8 (0.21)	0.81 (0.2)
Precision		
Coder 1	0.78 (0.21)	0.79 (0.23)
Coder 2	0.86 (0.17)	0.82 (0.18)
Coder 3	0.75 (0.18)	0.81 (0.19)
Coder 4	0.78 (0.23)	0.83 (0.2)
All coders	0.79 (0.2)	0.81 (0.2)
F-score		
Coder 1	0.79 (0.2)	0.81 (0.2)
Coder 2	0.84 (0.18)	0.8 (0.18)
Coder 3	0.78 (0.16)	0.82 (0.16)
Coder 4	0.73 (0.26)	0.78 (0.2)
All coders	0.79 (0.19)	0.80 (0.19)

Table 3

Effect of map-assisted coding according to coders' experience (* statistically significant)

Coder experience	Average change for map-assisted vs. usual coding			
	coding time (minute)	recall	precision	F-score
More experienced	-2.77*	0.0297	0.0297	0.0325
Less experienced	-0.18	-0.003	0.008	-0.0001

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4.

Reason of missing gold standard ICD-10-CM codes among the NLM Map suggestions

Reason for failure	Missing codes (%)	Examples			
		Problem term entered by physician	SNOMED CT term associated with problem term	Map-suggested ICD-10-CM code	ICD-10-CM code in gold standard
Problem term missing	63 (39%)	(No mention of obesity in problem list)	(none)	(none)	Obesity; unspecified (E66.9)
Problem term suboptimal	35 (21%)	Osteoarthritis of hip	Osteoarthritis of hip (239872002)	Osteoarthritis of hip, unspecified (M16.9)	Unilateral primary osteoarthritis; right hip (M16.11)
SNOMED CT term suboptimal	45 (28%)	Nevus	Skin lesion (95324001)	Disorder of the skin and subcutaneous tissue, unspecified (L98.9)	Nevus; non-neoplastic (I78.1)
ICD-10-CM map suboptimal	20 (12%)	Needs flu shot	Needs influenza immunization (185903001)	Underimmunization status (Z28.3)	Encounter for immunization (Z23)

Table 5

Raw and adjusted performance of map-suggested codes (the “NLM auto-coder”) vs. the average coder with usual coding.

Accuracy measure	“NLM auto-coder” accuracy according to		Average of all coders, usual coding, according to original gold standard
	Original gold standard	Adjusted gold standard	
Recall	0.71	0.95	0.8
Precision	0.63	0.73	0.79
F-score	0.66	0.81	0.79

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript