The ICD-11 Field Trial: Creating a Large Dually Coded Database

Cathy A Eastwood (caeastwo@ucalgary.ca)  
University of Calgary  https://orcid.org/0000-0002-4569-8014

Danielle A Southern  
University of Calgary Department of Community Health Sciences

Shahreen Khair  
University of Calgary  https://orcid.org/0000-0002-1685-0603

Chelsea Doktorchik  
University of Calgary Department of Community Health Sciences

William A Ghali  
University of Calgary

Hude Quan  
University of Calgary Department of Community Health Sciences

Research note

**Keywords:** Dually Coded Database, ICD-10-CA, ICD-11, Chart Review, Inter-rater reliability, Data, ICD-11 Field Trial, ICD-11 Reference Guide

**DOI:** https://doi.org/10.21203/rs.3.rs-505934/v1

**License:** ☝️ 📧 This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Objective

New codes developed in the International Classification of Diseases, 11th Revision for Mortality and Morbidity Statistics (ICD-11) needed testing. Field-testing involves real-world application of the new codes to examine data quality. This paper describes the field trial methods to create a dually coded database to field test ICD-11 against ICD-10-CA (a Canadian modification of the International Classification of Diseases, Tenth Revision), and a reference standard data set of diagnoses.

Results

A random sample of discharge records previously coded using ICD-10-CA was selected. Nurses re-examined these entire charts for specific conditions and patient safety events. Clinical coders re-coded the same charts using ICD-11 codes.

Inpatients discharged from hospitals in Calgary, Alberta, were identified and a dually coded database was created (n=2897). Inter-rater reliability and coding time improved with ICD-11 coding experience. Clinical coder comments enabled content to be improved in the ICD-11 browser, Coding Tool, and ICD-11 Reference Guide. This paper describes the field trial, database creation methods, and contributions for ICD-11 improvement. Crucial future research will use this database to test ICD-11 before implementation in Canada.

Introduction

Coded health data are important for many applications, including health services funding, physician payment, and research (1). Quality and usability of updated International Classification of Diseases (ICD) codes need to be examined before use.

The World Health Organization (WHO) introduced the Beta Version of ICD for Mortality and Morbidity Statistics, 11th Revision (ICD-11) for public discussion in 2012. New features include: 1) flexible code-clustering allowing rich descriptions of complex clinical scenarios; 2) new extension code chapter for enhanced detail on disease severity, progression, and timing; and 3) digital ICD-11 browser and coding tool to facilitate searching and linking to electronic health records (EHRs) (1,2). The WHO encouraged systematic testing of ICD-11 before its release to the public in June 2018.

In 2014, the WHO designated the University of Calgary as an international academic Collaborating Centre (WHO-CC). This WHO-CC undertook a large-scale field trial that served to inform further ICD-11 development through real-world coding.

This report outlines the methods used to test ICD-11 using a dually coded database using both ICD-10-CA (Canadian modification) and ICD-11 codes, and chart review as a reference. A database with gold standard labelled records is essential to validate code accuracy or when data mining for specified
conditions. Without condition labels, validity cannot be calculated. This database will be used for future studies involving mapping codes, comparability across ICD versions, improving codes and definitions for ICD-11, and trend analysis of diseases.

Materials And Methods

We generated and linked three data sets: 1) a retrospective clinical chart review as a reference standard; 2) original ICD-10-CA coded data; and 3) re-coded ICD-11 coded data (Figure 1). To date, data collection is complete, and analyses are underway.

Sample Size and Cohort

Based on previous (3) findings on sensitivity and prevalence of conditions in a sample of ICD-10-CA data, 3000 records were deemed required to test a 10% difference in sensitivity of common conditions such as myocardial infarction (12.8%), cardiac arrhythmia (21.8%), hypertension (30.2%), and others. The Lachenbruch's (4) midpoint method was used.

The study cohort included a random sample of discharges selected from records between January 1st and June 30th of 2015, from three hospitals in Calgary, Alberta. Patients were 18 and 104 years of age with a valid Personal Health Number (PHN) for Alberta. Obstetric admissions were excluded due to short stay and absence of chronic conditions of interest. The first 1100 records from each hospital with the lowest random chart numbers were selected. If there were multiple discharges for a single patient during the study period, we randomly selected one discharge record per patient. The additional 100 records per site allowed for missing or excluded charts.

Chart Review Dataset

Internal validation of a dually coded database involves measuring how well codes, selected from both ICD-10-CA and ICD-11, represent the diagnoses identified by chart reviewers, in terms of sensitivity, specificity, positive and negative predictive values.

Data Dictionary

We replicated and expanded the chart review approach from our prior study on the validity of ICD-10-CA (3). We selected 51 medical conditions, including the Charlson and Elixhauser (5,6) comorbidity conditions, and up to three harms (Table 1). We chose these conditions from other validation studies (3,7) and they are commonly used for risk adjustment. Some definitions were based on literature (5,6) and our prior validation study (3). Where no published definition was available, ICD-11 Browser definitions (beta version) were used (2). Chart review conditions are listed in Table 1.

Table 1 Chart Review Conditions for Data Collection
| Conditions                         | 1. Angina                   | 18. Inflammatory bowel disease | 35. HIV/AIDS         |
|-----------------------------------|-----------------------------|--------------------------------|----------------------|
| 2. Myocardial infarction (new)    | 19. Liver disease           | 36. Disorders due to tobacco use |
| 3. Myocardial infarction (old)    | 20. Cancer                  | 37. Dyslipidemia                |
| 4. Congestive heart failure       | 21. Malignancy without metastases | 38. Disorders due to alcohol use |
| 5. Cardiac arrhythmias            | 22. Malignancy with metastases | 39. Disorders due to drug use    |
| 6. Atrial fibrillation            | 23. Leukemia                | 40. Psychoses                   |
| 7. Atrial flutter                 | 24. Lymphoma                | 41. Anxiety                     |
| 8. Valve disease                  | 25. Renal disease           | 42. Depression                  |
| 9. Pulmonary circulatory disorders| 26. Rheumatologic disease   | 43. Homeless                    |
| 10. Hypertension                  | 27. Diabetes                | 44. Urinary tract infection     |
| 11. Peripheral vascular disease   | 28. Hypothyroidism          | 45. Pneumonia                   |
| 12. Cerebrovascular disease       | 29. Coagulopathy            | 46. Skin/wound infection        |
| 13. Paralysis                     | 30. Anemia                  | 47. Gastroenteritis             |
| 14. Chronic pulmonary disease     | 31. Fluid & electrolyte disorder | 48. Other infection            |
| 15. Asthma                        | 32. Obesity                 | 49. Sepsis                      |
| 16. Peptic ulcer Disease          | 33. Significant weight loss | 50. Pressure Ulcer              |
| 17. Gastrointestinal bleed        | 34. Dementia                | 51. Sleep disorders             |

**Harms**

| Harms  | Harm 1 | Harm 2 | Harm 3 |
|--------|--------|--------|--------|

Detailed condition definitions are available in the Data Dictionary for ICD-11 Field Trial [see Additional file 1].

Chart Access
Patient charts were available in paper and electronic (hybrid) form in each hospital's health records department. Electronic content was accessed in Sunrise Clinical Manager™ (SCM).

Chart Review Team

Six nurse chart reviewers underwent extensive training on the data extraction process by the research coordinator. Training involved learning the data dictionary definitions and following a consistent order to review the chart documents. To test the data definitions, training included practice identifying the medical conditions in the same five charts. Discrepancies between the reviewers were discussed and the data dictionary was refined. We then proceeded with inter-rater reliability (IRR) explained below.

The nurse reviewers examined the entire chart for the presence of specific health conditions. These reviewers were blinded to the ICD codes assigned by the coders.

ICD-10-CA Coded Dataset

We used previously coded charts because the existing ICD-10-CA dataset represented a “real-life” sample of coding practices. Alberta hospitals employ trained clinical coders (CCs) (i.e., nationally certified health information management specialists) who read through patient hospital charts. These CCs assigned ICD-10-CA diagnosis codes to describe each patient's hospitalization, based on ICD-10-CA Canadian coding standards (8). Each discharge record contains a unique identification number for each admission and up to 25 fields for diagnosis codes, which became the study dataset.

Re-coded ICD-11 Dataset

The third phase involved re-coding the same inpatient charts using ICD-11.

Training Material Development

The research coordinator and a member of Canadian Institute developed ICD-11 training materials for Health Information (CIHI). Materials included three slide sets covering ICD-11 concepts and tools (9). Materials for coding practice included two sets each of Morbidity and Quality and Safety Case Scenarios. Coding rules and decision trees were developed for coding hospital-acquired conditions (harms) in conjunction with the WHO Quality and Safety Technical Advisory Group (TAG). The full training process is presented in Eastwood et al. (submitted) (10).

Clinical Coding Team

Six professional CCs were hired and trained. Trainers included a team from University of Calgary, CIHI, and WHO experts in ICD-11 concepts. Training involved 20 classroom hours and approximately 40 hours of coding practice homework prior to coding full hospital charts. Then, the coding team and trainers met
monthly during the coding phase to discuss coding issues. ICD-11 coding decisions were based on what was available at the time in the draft ICD-11 Reference Guide of the WHO (11), the WHO ICD-11 Coding Tool (12), and the Canadian ICD-10-CA coding standards (8), given that ICD-11 coding rules were limited.

Analysis

Test Inter-rater Reliability (IRR) of Chart Review

To test agreement between reviewers, IRR involved two nurses reviewing sets of the same 10 charts. Agreement was checked for the presence of the 17 Charlson conditions. Where agreement was poor (kappa<0.60) retraining took place and chart review resumed in batches of 10 charts, until agreement was high (kappa>0.8) (13). High agreement was reached after two people completed 49 sets of records. Reviewers then independently extracted data from the remaining charts over several months. Data were entered into a secure electronic data collection tool called REDCap (7.6.9-©2018 Vanderbilt University). IRR was not available for the previously coded ICD-10-CA dataset.

Test Inter-rater Reliability of ICD-11 Coded Charts

IRR involved 60 full charts coded by two CCs, similar to the above chart review IRR process. IRR focused on consistent coding of the main condition given the bulk of possible codes generated from full hospital charts. After the first 40 charts, a kappa of 0.50 was reached on the main condition parent code (highest level in the ICD-11 condition hierarchy). Training continued, differences were discussed, and experts were engaged for guidance. After coding the next 20 charts, a kappa of 0.88 was reached for main condition parent codes and independent coding commenced. The CCs were blinded to the original ICD-10-CA codes and the chart review data.

Results

This paper describes the methods for creating a dually coded database, to test ICD-11. Results include the final number in the database, coding time, and recommendations made to WHO to improve the ICD-11 browser.

The chart review sample included 3045 records. The sample of charts coded with both ICD-10-CA and ICD-11 was 3011. The final sample for the dually coded database, with complete data for all three data sets, was n=2897. Complete meant that all chart review fields were filled, and coded charts had at least one diagnosis code.

Chart review time per chart averaged 14.6 minutes (std. dev. = 29.1), median 11.0 minutes. Coding time using ICD-11 declined from 23.6 minutes (std. dev. =14.1) to 9.9 minutes (std. dev. = 6.4) on average per chart, as coding proficiency increased during the study (Figure 2).
Recommendations made for the ICD-11 Reference Guide and ICD-11

This study enabled feedback to the WHO on the new codes and Reference Guide. Changes were integrated into the ICD-11 Browser prior to its release in June 2018. Changes to the ICD-11 Reference Guide for the morbidity-related chapters included improved clinical definitions, and expanded instructions on cluster coding and post-coordination (11). Substantial content was added to clarify Chapter 23, External Causes of Morbidity and Mortality. As such, ICD-11 Reference Guide now includes a framework and guidelines for using the three-part model to code healthcare-related harms (11). ICD-11 improvements included resolving missing codes and inclusion terms, post-coordination linkages, codes choices when documentation was ambiguous, substance/medication list, coding harms with the 3-part model, and features and functions of the Coding Tool. Example ICD-11 changes are listed [see Additional file 2].

Discussion

This detailed set of methods is available for use in other countries testing and adopting ICD-11. Our coding team was the first to code a large number of full hospital discharge records with ICD-11. Developing this dually coded database greatly contributed to refining the new classification system for all stakeholders to benefit.

Learning Points: ICD-11 Coding Training

Challenges with inter-rater reliability for ICD-11 coding were multifactorial. The comprehensive ICD-11 contains 55,000 unique codes (14), thus more code choices, while ICD-10-CA contains only 12,420 codes (15). Also, at the time of data collection, codes and coding procedures for ICD-11 were still under revision, making training and learning challenging. The CCs required training for the new code structures like code clustering (16). While information sheets were referenced from WHO Education and Implementation Committee (EIC) (17), more training materials needed to be developed by our team. Even with these resources, systematic training for coding specialists on the coding of complex case scenarios was challenging. Of note, new ICD-11 training materials are now available from the WHO (18) and the EIC committee (17), and the ICD-11 Browser, Reference Guide, and Coding Tool are refined (2).

Learning Points: Chart Review Data Collection Quality

Training clinical chart reviewers is crucial for reliable reference data. For optimal data collection, we ensured chart reviewers had the clearest possible data definitions and a set of steps to systematically locate conditions within the chart. The team often collected data in a collaborative environment to resolve discrepancies in real-time. A significant contribution of this study is our detailed data definition dictionary, given minimal definitions were previously published for the Charlson and Elixhauser conditions (5,6). These definitions offer potential standard definitions for future research.
Incomplete chart documentation was a real-world problem during coding and nurse chart review, which caused unavoidable gaps in data collection. We anticipate that documentation quality will improve as our health system transitions to electronic records, potentially enhancing chart completeness.

**Conclusion**

This paper describes the methods for creating a dually coded ICD-11 and ICD-10-CA database. The study was timely and provided recommendations for ICD-11 enhancement prior to its public release. These methods can be replicated for other code detection and validation studies worldwide. Future studies using this dually coded database and reference standard will examine ICD-11 code features, as well as ICD-11 coded data validity for common clinical conditions.

**Limitations**

Initially, IRR for ICD-11 coding was low between CCs but improved with discussion and re-training. Common reasons for low IRR may include 1) more codes to choose from in ICD-11 and 2) limited formal guidelines and reference materials. Re-testing IRR and retraining at various intervals would further strengthen IRR. While ICD-10-CA codes were collected in a “real-life” setting with various CCs, ICD-11 codes were collected in a controlled research setting with a small pool of trained CCs. Previously coded ICD-10-CA data were chosen to reduce resource use in the current study and to reflect typically collected coded data.

**Abbreviations**

**ICD**: International Classification of Diseases  
**ICD-11**: International Classification of Diseases, 11th Revision for Mortality and Morbidity Statistics  
**ICD-10-CA**: Canadian specific version of the International Classification of Diseases, Tenth Revision  
**CC**: Clinical Coders  
**WHO**: World Health Organization  
**EHR**: Electronic Health Record  
**WHO-CC**: World Health Organization Collaborating Centre  
**CHREB**: Conjoint Health Research Ethics Board  
**PHN**: Personal Health Number  
**SCM**: Sunrise Clinical Manager™
Declarations

Ethics approval and consent to participate

Ethics approval was obtained from the Conjoint Health Research Ethics Board from the University of Calgary (REB15-0790). A waiver of consent was granted for retrospective chart review and administrative data extraction. A Research Agreement was signed with Alberta Health Services Analytics who is the data custodian.

Consent for publication

Not Applicable

Availability of data and material

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the Canadian Institutes of Health Research [grant ID 142182].

Author’s contributions

CE conceptualized the study, interpreted the data and contributed to manuscript writing. DS contributed to conceptualising the analysis and manuscript writing. SK and CD contributed to manuscript writing and analysis. WG and HQ conceptualized the study and contributed to funding acquisition and analysis. All authors critically revised the paper and edited and approved the final manuscript.
Acknowledgements

We would like to thank the Centre for Health Informatics at the University of Calgary for their support on this project. We acknowledge the excellent work from the chart reviewers: Olga Grosu, Chris King, Natalie Wiebe, Ellena Kim, Danielle Fox, and Nicholas Van Kampen. Further, we would like to thank the CCs on our team: Alicia Boxill, Tia Schafhauser, Natasha Hanus, Jennie Ma, Jessica Gillis, Clair Selebay, and Nicky (Pallavi) Mathur. We acknowledge the ongoing support from our collaborators at CIHI, Denise Cullen and Margaret Penchoff. Finally, we would like to thank Lori Moskal at the WHO for her support and guidance on this project.

References

1. World Health Organization. ICD – 11 Fact Sheet. 2019;1–3. Available from: https://icd.who.int/en/docs/icd11factsheet_en.pdf
2. World Health Organization. ICD-11 for Mortality and Morbidity Statistics [Internet]. 2020. Available from: https://icd.who.int/browse11/l-m/en
3. Quan H, Li B, Duncan Saunders L, Parsons GA, Nilsson CI, Alibhai A, et al. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. Health Serv Res. 2008;43(4):1424–41.
4. Lachenbruch PA. On the sample size for studies based upon mcnemar’s test. Stat Med. 1992;11(11):1521–5.
5. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity Measures for Use with Administrative Data. Med Care [Internet]. 1998;36(1):8–27. Available from: http://www.jstor.org/stable/3766985
6. Charlson ME, Pompei P, Ales KL, MacKenzie C. A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation. J Chronic Dis [Internet]. 1987;40(5):373–83. Available from: http://www.sciencedirect.com/science/article/pii/0021968187901718
7. Doktorchik C, Patten S, Eastwood C, Peng M, Chen G, Beck CA, et al. Validation of a case definition for depression in administrative data against primary chart data as a reference standard 11 Medical and Health Sciences 1117 Public Health and Health Services. BMC Psychiatry. 2019;19(1):1–8.
8. Canadian Institute for Health Information. Canadian Coding Standards for Version 2018 ICD-10-CA and CCI. Ottawa, ON; 2018.
9. Eastwood CA, Southern DA, Doktorchik C, Khair S, Cullen D, Boxill A, et al. Foundational concepts of ICD-11 with case scenarios and coding example worksheets to supplement learning and implementation. [Internet]. [cited 2020 Sep 27]. Available from: https://cumming.ucalgary.ca/node/409627
10. Eastwood CA, Southern DA, Doktorchik C, Khair S, Cullen D, Boxill A, et al. Training and Experience of Coding with WHO’s International Classification of Diseases 11th Revision (ICD-11).
11. World Health Organization. ICD-11 Reference Guide [Internet]. 2019 [cited 2020 Sep 27]. Available from: https://icd.who.int/icd11refguide/en/index.html#2.25.5CodesetPSQ%7Ccausation-in-the-context-of-quality-and-safety%7Cc2-25-5-1
12. ICD-11 Coding Tool [Internet]. 2019. Available from: https://icd.who.int/ct11/icd11_mms/en/release
13. Kunac DL, Reith DM, Kennedy J, Austin NC, Williams SM. Inter- and intra-rater reliability for classification of medication related events in paediatric inpatients. Qual Saf Heal Care. 2006;15(3):196–201.
14. Lindmeier C. WHO releases new International Classification of Diseases (ICD 11). 2018; Available from: https://www.who.int/news/item/18-06-2018-who-releases-new-international-classification-of-diseases-(icd-11)#:~:text=18 June 2018 ¦ Geneva,diseases and causes of death
15. World Health Organization. FAQ on ICD [Internet]. 2020 [cited 2020 Dec 2]. Available from: https://www.who.int/classifications/help/icdfaq/en/
16. The American Health Information Management Association. AHIMA [Internet]. 2018 [cited 2019 Apr 18]. Available from: http://bok.ahima.org/doc?oid=59978#.XLYSMi3MxQJ).
17. NHS Digital. WHO-FIC Education & Implementation Committee, Resource Library [Internet]. 2020 [cited 2020 Nov 7]. Available from: https://hscic.kahootz.com/connect.ti/WHO_FIC_EIC/viewdatastore?dsid=521828&adv=S&showAllColumns=N&datViewMode=list&showSingleItem=N&cardColNo=
18. World Health Organization. ICD-11 Implementation or Transition Guide. ICD-11 Implement or Transit Guid [Internet]. 2019;1–29. Available from: https://icd.who.int/docs/ICD-11 Implementation or Transition Guide_v105.pdf

Figures
Figure 1

Steps for creating a chart review reference dataset, and a dually coded dataset
Figure 2

Coding Over Time with ICD-11

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile1ofDataDictionaryforICD11FieldTrial.docx
- AdditionalFile2ofExamplesofchangesmadetoICD11.docx