Piloting a Collaborative Web-Based System for Testing ICD-11

Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-466

Piloting a Collaborative Web-Based System for Testing ICD-11

Marc DONADA[a], Nenad KOSTANJSEK[b] ,Vincenzo DELLA MEA[a,1], Can CELIK[b], Robert JAKOB[b]
[a]*Dept. of Mathematics,Computer Science and Physics, University of Udine, Italy*
[b]*Classifications,Terminologies and Standards, World Health Organization, Geneva, Switzerland*

Abstract. Background: The 11th revision of the International Classification of Diseases (ICD-11), for the first time in ICD history, deployed web-based collaboration of experts and ICT tools. To ensure that ICD-11 is working well, it needs to be systematically field tested in different settings, across the world. This will be done by means of a number of experiments. In order to support its implementation, a web-based system (ICDfit) has been designed and developed. The present paper illustrates the current prototype of the system and its technical testing. Methods: the system has been designed according to WHO requirements, and implemented using PHP and MySQL. Then, a preliminary technical test has been designed and run in January 2016, involving 8 users. They had to carry out double coding, that is, coding case summaries with both ICD-10 and ICD-11, and answering quick questions on the coding difficulty. Results: the 8 users coded 632 cases each, spending an average of 163 seconds per case. While we found an issue in the mechanism used to record coding times, no further issues were found. Conclusion: the proposed system seems to be technically adequate for supporting future ICD-11 testing.

Keywords. International Classification of Diseases, Clinical Coding, Questionnaires

1. Introduction

The 11th revision of the International Classification of Diseases (ICD-11) has been recently presented to Member States for comment [1,2]. For the first time in the history of ICD revisions web-based expert collaboration and ICT tools were used in the development of the classification [3]. The ongoing ICD revision process is responding to numerous demands to align the classification to the latest scientific evidence and user requirements for more purpose-driven and IT compatible capturing and processing of diagnostic information.

To ensure that ICD-11 is working well, it needs to be systematically tested in different settings, across the world. This will be done by means of a number of testing arrangements.

The ICD-11 testing is unprecedented in the ICD revision history as in previous ICD revisions field testing was limited in scope (i.e. international field testing ICD-10

[1] Corresponding Author, Department of Mathematics, Computer Science and Physics, University of Udine, via delle Scienze 206, 33100 Udine, Italy. Email: vincenzo.dellamea@uniud.it

Chapter V) or conducted as an after step to facilitate the transition between the old and the new classification system (i.e. national Bridge Coding studies between ICD-9 and ICD-10) [4,5].

The overarching objective of the ICD-11 testing is to ensure systematic testing of ICD-11 before its use to increase consistency, accuracy and usability for morbidity and mortality coding as well as ascertain the comparability between ICD-10 and ICD-11.

To warrant strong and organized participation in the field testing from around the world WHO will designate ICD-11 Field Test Centers (FTC), which in turn will manage a network of national sites participating in the test. WHO Collaborating Centers, ICD-11 Topic Advisory Groups or other organizations with sufficient implementation capacity can serve as FTC.

In order to support the testing, a web-based system (ICDfit) has been designed and developed. To ensure functionality with large number of users an ICDfit prototype is currently undergoing a pilot testing. The paper describes features of the current ICDfit prototype and the process of its pilot testing.

2. System Requirements and Design

The system should implement the organization structure of field tests (FT) as foreseen by WHO, based on WHO coordination, FT centers (FTC) and FT sites (FTS). Two main kinds of studies are foreseen: one is devoted to line and case coding for mortality and morbidity using ICD-11 codes, or using both ICD-10 and ICD-11 codes. The diagnostic terms or statements used for line or case coding are extracted from clinical records, death certificates, etc. Another kind of study is devoted to asking basic questions on the updated classification to stakeholders, i.e., policy makers, classification experts, etc.

Users can belong to different categories: FTC Coordinator, FTS coordinator, Rater and Key Informant. Users have the following roles:

- The **Rater** carries out the basic work of participating into coding studies as classification user (i.e., coder). He/she accesses personalized ICD-FiT web page and checks for assigned cases in field test studies, completes case related forms for the respective study**,** fills in the evaluation form after completing all cases of a study.
- The **Key informant** participates into basic questions studies as classifications stakeholder. He/she accesses personalized ICD-FiT web page and fills in the forms related to the basic questions.
- An **FTS Coordinator** is able to invite and manage raters, and assign cases to them for each study in which the site is involved.
- An **FTC Coordinator** has the ability to assign new field test sites in their country, invite and assign key informants for basic questions studies, and assign cases to raters for each study in which the site is involved. This user role also has responsibility of translation of case summaries in local languages.
- The **WHO Coordinator** is able to insert new field test centers in the system, assign them to field test studies as well as inserting cases.

Collected data should be exported in CSV and Excel format for further processing.

Finally, the system should be sufficiently flexible to be adapted to new studies if needed.

Tools for web-based data collection exist, ranging from simple survey tools (like SurveyMonkey, Google Forms, etc) to complex systems addressing data collection in clinical trials (e.g., OpenClinica, REDCap, etc) [6]. However, due to the specific hierarchical organization of the field testing, together with the multilingual support needed also at the level of case summaries, we decided to implement an *ad-hoc* system specialized for the above mentioned requirements.

The system has been developed using PHP5, MySQL, Apache, on a Linux server. The three main FT instruments have been developed in both Western and Traditional Medicine version. An additional variant of the coding instrument, called "line coding", has been implemented to facilitate quick response.

Multilingual support has been developed, so that the system is currently available in 6 languages (although cases are only partially translated). The web interface has been developed using a responsive template in order to be accessible from tablets and smartphones too. Figure 1 shows an example of the web interface for double coding: on top there is a short case summary that has to be coded by raters, then there is room for up to 3 ICD-11 codes and the survey questions. After the end of ICD-11 coding, clicking on the "Start ICD-10 coding" will reveal an identical form section for ICD-10.

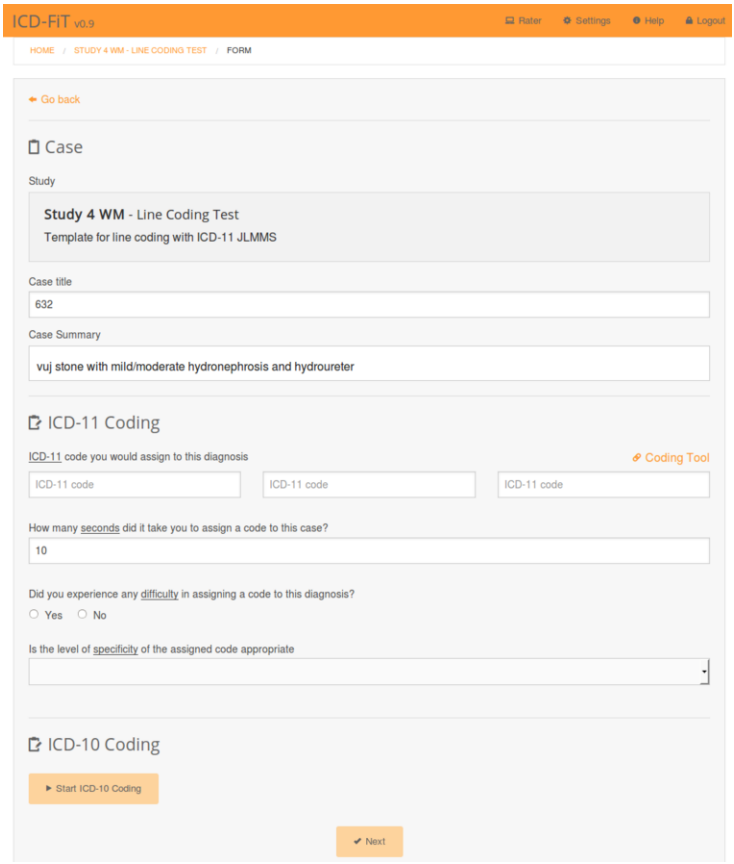

Figure 1. Line coding form.

3. Preliminary Technical Testing

In order to verify the functionality of the system, with also some preliminary indication on time needed for its usage and possible issues in the implementation, a technical test has been designed with the following features:

- 632 case summaries in English language have been prepared by the WHO coordinator and inserted into the system;
- A line coding instrument has been developed inside ICDfit to allow double coding using ICD-10 and ICD-11. Up to three codes per case summary were allowed for both ICD-10 and ICD-11 (fig.1);
- 8 coders from the Korean WHO-FIC collaboration center have been involved in coding the case summaries;
- The tool suggested for selecting codes in ICD-10 was the official ICD-10 browser, while for ICD-11 the new, experimental coding tool has been selected;
- Users were asked to code first using ICD-11, then using ICD-10.

Data collected included codes (up to 3 per case and per classification), answers to 3 questions per classification (regarding coding difficulties and code specificity), and time information (total time per case, time for ICD-10 coding, time for ICD-11 coding). A separate log of actions carried out on the system has also been maintained, in order to investigate on possible errors and issues.

The system has been integrated with available tools developed at WHO, in particular the ICD-10 browser and the ICD-11 coding tool.

4. Results

The technical testing experiment has been run in January 2016 and was successfully completed in two weeks by all coders, for a total of 5056 coded case summaries.

All cases were coded with at least one code, 153 had a second code in ICD-11 and 96 in ICD-10 by at least one coder, only 12 had a third code in ICD-11 vs. 6 in ICD-10.

The average time needed for double-coding a case and answering the survey questions was 163 seconds (range: 14-1426). However, when looking at the individual times for coding with just one classification, we found some discrepancies in the time spent. In fact, the average time for ICD-11 was 38 seconds, and for ICD-10 was 29 seconds.

An interview with users that revealed these discrepancies allowed to find that they were using a coding method that was more effective to them, but circumvented the time measuring method adopted in the system. In fact, some users maintained 3 browser tabs open with ICDFit, ICD-10 browser and ICD-11 coding tool: after having read the case summary, they individuated both codes on the respective tools, and then pasted them into the line coding tool.

Regarding the survey questions, coding difficulty was slightly more for ICD-11, as well as the number of answers indicating too much or too broad detail in ICD-11 codes. This is partly due to the not yet completed status of ICD-11. However, it should be noted that ICD-11 evaluation was not the aim of the test, more oriented towards technical functionalities of the system.

A thorough examination of server logs revealed relatively few problems. Some of them were related to network issues at very specific times; some other to session timeout due to inactivity or unstable network, that caused users to login again to complete their work. Session length was set at 24 minutes.

5. Discussion and Conclusions

The users participating in the technical testing were able to complete the assigned case summaries coding exercise in a relatively short time. The experiment allowed to discover some issues in the automatic time computation, due to the different behavior patterns users had in coding. This way, the apparent time needed for coding, as measured by the input timeline, was really short, although the total time resulting from logs was correct. Since it is almost impossible to technically constrain a more sequential behavior, if there is a real need for a precise evaluation of coding time, strict instructions should be provided to raters for the participation in the study.

However, the shortcuts identified by raters also suggest ways to make the overall field test process quicker, which is by itself a good result.

The integration with WHO tools has been also proven to function well, in particular for the newly released ICD-11 coding tool, that allows for natural language queries.

The overall time needed to double code a case is relatively short and sustainable, even in view of the tenths thousands answers needed for field test. However, before fully testing ICD-11, smaller tests will be made on selected, crucial chapters of the classification.

ICDfit has been adequate for helping to document and measure the code assignment in ICD-11 and ICD-10. Due to its generic features, it could also be used for other classifications. Further work will include mainly the development and implementation of a metrics to compare raters coding with gold standard coding, based on their distance on the classification tree, to be used for both ICD-10 and ICD-11.

References

[1] World Health Organization, “The International Classification of Diseases 11th Revision is due by 2018”, http://www.who.int/classifications/icd/revision/en/ (last accessed 10 Mar 2016).

[2] Boerma T, Harrison J, Jakob R, Mathers C, Schmider A, Weber S. Revising the ICD: explaining the WHO approach. Lancet. 2016 Oct 14.

[3] T.Tudorache, S.Falconer, C.Nyulas, M.A.Storey, T.B.Ustün, M.A.Musen, “Supporting the Collaborative Authoring of ICD-11 with WebProtégé”, Proc. of AMIA Annu Fall Symp. 2010: 802-6.

[4] Stieglitz RD, Zaudig M, Freyberger HJ, Dittmann V. Feasibility, suitability, and interrater reliability of ICD-10 during different stages of the ICD-10 field trial. Pharmacopsychiatry. 1990;23 Suppl 4:188-91

[5] Ustün TB, Goldberg D, Cooper J, Simon GE, Sartorius N. New classification for mental disorders with management guidelines for use in primary care: ICD-10 PHC chapter five. Br J Gen Pract. 1995 Apr;45(393):211-5.

[6] Franklin JD, Guidry A, Brinkley JF. A partnership approach for Electronic Data Capture in small-scale clinical trials. J Biomed Inform. 2011 Dec;44 Suppl 1:S103-8.