



MEETING OF HEADS OF WHO COLLABORATING CENTRES FOR THE CLASSIFICATION OF DISEASES

Brisbane, Queensland, Australia
14-19th October 2002

Title: Consistency Checks for the Maintenance of ICD-10

Author: Michael Schopen, DIMDI

Purpose: for information

Abstract:

This paper describes certain routines implemented in the SGML based production system for the German language edition of ICD-10. They serve to enforce consistency of the master files and thus consistency of the user files. Checks for the following purposes will be presented:

- Automatic generation of certain parts of the Tabular List (redundant information)
- Checks to Volume 1
- Checks to Volume 3
- Cross-checks between Volume 1, Volume 2 and Volume 3

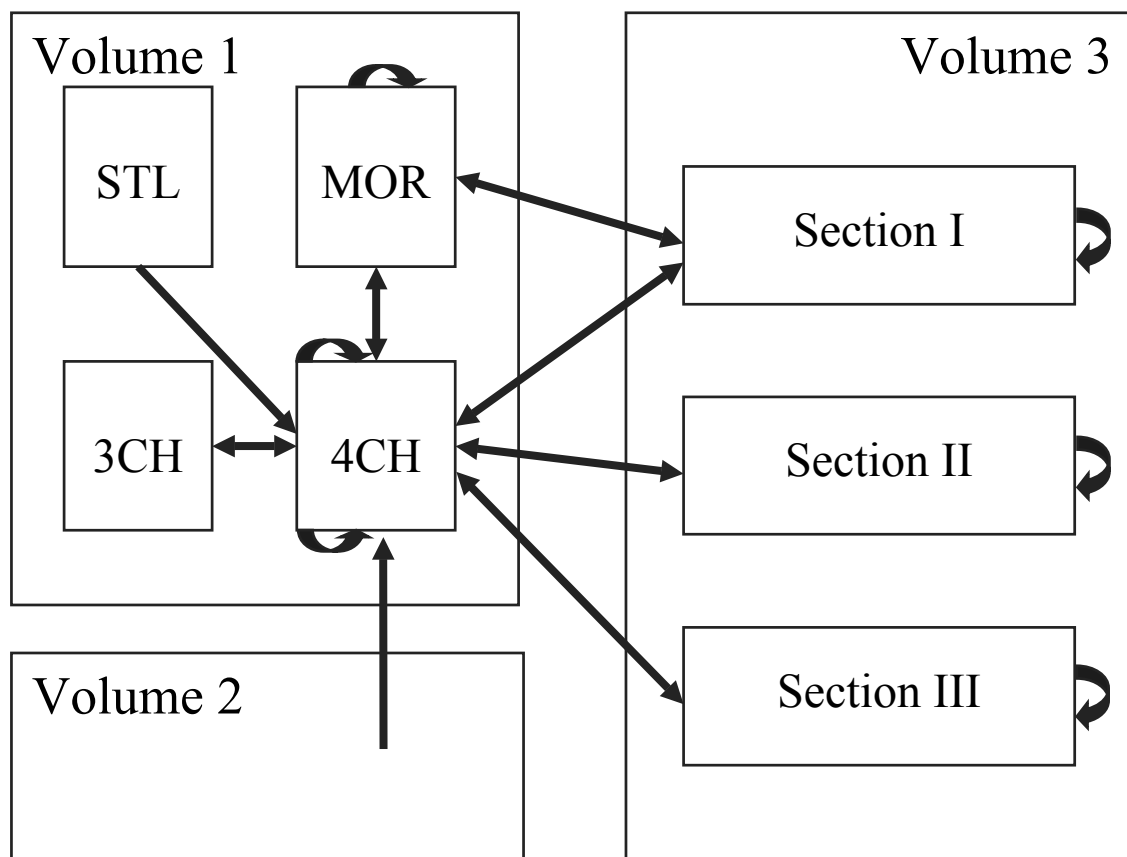
This document is not issued to the general public, and all rights are reserved by the World Health Organization (WHO). The document may not be reviewed, abstracted, quoted, reproduced or translated, in part or in whole, without the prior written permission of WHO. No part of this document may be stored in a retrieval system or transmitted in any form or by any means - electronic, mechanical or other - without the prior written permission of WHO.

The views expressed in documents by named authors are solely the responsibility of those authors.

Introduction

Creation and maintenance of a classification is an error-prone process. Typographic errors are only a minor problem, although they can prohibit computerized applications considerably, as searches might produce zero hits when the user enters search terms in correct typography. However, inconsistencies in the classification create major problems as they lead to ambiguities during the coding process which are a major obstacle to data comparison.

The following figure displays the dependencies between Volume 1, Volume 2 and Volume 3 of ICD-10:



The relevant parts from Volume 1 are the "Tabular List of inclusions and four-character subcategories" (abbreviated 4CH), the "List of three-character categories" (3CH), the "Morphology of neoplasms" (MOR) and the "Special tabulation lists for mortality and morbidity" (STL). Volume 3 consists of three sections: The "Alphabetical index to diseases and nature of injury" (Section I), the "External causes of injury" (Section II) and the "Table of drugs and chemicals" (Section III). There are dependencies within Volume 1 - e.g. between 3CH and 4CH - and dependencies within certain parts of Volume 1 - e.g. within 4CH or MOR. Furthermore, there are dependencies between Volume 1 and Volume 3 - e.g. between Sect I and 4CH - and between Volume 1 and Volume 2 - e.g. examples with codes from 4CH or lists with coding instructions for codes from 4CH. These dependencies make updates to ICD-10 an error-prone process. Thus, computerized assistance is necessary to reduce errors as far as possible.

This paper describes consistency checks for the maintenance of ICD-10 and is to continue earlier papers on the publishing of the German language edition of ICD-10 (1, 2). Constraints are presented together with some general considerations on their implementation. Many of them have now been implemented in the SGML-based maintenance system for the German language edition of ICD-10 at DIMDI.

Automatic generation of redundant information

When a three-character category is added to or deleted from the Tabular List or when its title is modified, then these changes will necessarily affect not only 4CH but also 3CH. Some changes might also affect the *coding frame* of the relevant chapter and block, the *list of blocks* or the *list of asterisk codes* at the beginning of the chapter. This information is redundant, but makes book reading easier or may serve as navigation aids in computerized versions. As redundancies are always a source of errors, this information should not be duplicated in the maintenance files.

Constraint T1: Redundancy needed for the final version should not be achieved by data duplication in the maintenance files.

Implementation: 3CH, the lists of blocks and asterisk codes, and the coding frames should be generated automatically from 4CH.

Consistency checks within single parts of Volume 1

4CH makes extensive use of references in the exclusion notes and the dagger and asterisk system. Also MOR uses a few exclusion notes. Errors during data entry might reference non-existing codes or - even worse - existing but wrong codes. The deletion of codes from the hierarchy might also leave references to non-existing codes.

Constraint T2: A reference from an exclusion note must point to existing codes or code ranges in the hierarchy.

Constraint T3: References within the dagger and asterisk system must point to existing nodes in the hierarchy.

Implementation: Online validation of all codes in the exclusion notes and in the dagger-asterisk system should be achieved during data entry. However, a final electronic check prior to production of the user files will also guarantee constraints T2 and T3.

Constraint T4: If an exclusion note repeats the title of an existing code elsewhere in the hierarchy, the texts should be identical.

Implementation: Again, online checks of the texts during data entry should be achieved. However, a final check prior to publication will also guarantee constraint T4. T4 will mainly be enforced manually. However, once consistency has been achieved, checks are only needed for those codes which are affected by an update.

Constraint T5: If a code is split into two or more "new" codes during an update, all exclusion notes referencing the "old" code must be checked.

Implementation: Similar to T4.

Consistency checks between the Tabular list of four-character categories and the Morphology of neoplasms or the Special tabulation lists

There are cross-references between 4CH and MOR: chapter II lists morphology codes and MOR codes from chapter II.

Constraint T6: References between 4CH and MOR must point to existing codes.

Implementation: As in T2

The entries in the Special tabulation lists are defined by codes and code ranges pointing to 4CH. There are no back-references from 4CH to STL.

Constraint T7: References from STL to 4CH must point to existing codes.

Implementation: As in T2

Furthermore, when codes are added to or deleted from the Tabular List, the codes and code ranges in the Special Tabulation Lists may need modification.

Consistency checks for Volume 3

The situation is easier than for Volume 1 as there are no references between the three sections. However, the *see* and *see also* references must be kept consistent.

Constraint A1: *See* and *see also* references in a section of the Alphabetical Index must point to existing entries and the texts of the reference and of the referenced entry should be identical.

Implementation: The implementation is quite tedious as each reference must be linked intellectually to the appropriate index entry. Afterwards, maintenance of these links is necessary during each update. For the German language edition, these links have meanwhile been established.

The three sections use different levels of indentation to lead the reader to certain entries. For reasons of computerization it is necessary to define which entries in the hierarchy of indentation are terminal (leaf nodes) and which are not. Terminal entries are entries with a code or a *see* reference. An entry must always be a terminal entry if the subsequent entry has the same level of indentation or a lower level of indentation.

Constraint A2: An entry, whose subsequent entry has the same or a lower level of indentation, must list at least one of the following data items: a single code or a dagger and an asterisk code from 4CH, a code from MOR, a *see* reference or a coding instruction ("code as ..."). Entries with a *see also* reference must either provide one of the data items listed above or they are non-terminal nodes.

Example:

Delirium, delirious (acute or subacute) (not alcohol- or drug-induced) F05.9

- due to (secondary to)
- - alcohol
- - - intoxication F10.0
- traumatic (*see also* Injury, intracranial)
- tremens (alcohol-induced) F10.4

An error occurs at "Delirium, traumatic", which is terminal in the hierarchy, as the subsequent entry has the same level of indentation, but it does not list one of the data items required by A2.

Implementation: In an SGML based system this constraint can already be enforced during data entry by the document type definition (DTD). The same can be achieved in data base systems.

Constraint A3: If a morphology code is accompanied by a code from chapter II, the behaviour code of this morphology code must be consistent with the corresponding block in chapter II.

Implementation: This constraint can easily be implemented by checking the behaviour code against the code ranges of the relevant blocks from the chapter II. E.g. morphology codes with dignity /6 can only be accompanied by codes from D10-D36.

Constraint A4: Dagger codes must be accompanied by asterisk codes and vice versa.

Implementation: This constraint can be enforced in a database or by the document type definition of an SGML document.

Consistency checks between Volume 1 and Volume 2

Constraint C1: Codes in Volume 2 must be valid nodes in the hierarchy of the Tabular List.

Constraint C2: If a code in Volume 2 is accompanied by its title, the texts in Volume 2 and in the Tabular List should be identical.

Constraint C3: If a code in the Tabular List is split into two or more "new" codes during an update, all references to the "old" code in Volume 2 must be checked.

Implementation: To ensure consistency, all codes in Volume 2 should be linked to the Tabular List. This can be achieved more or less automatically by converting them into hyperlinks. Once consistency has been achieved, only codes subject to an update must be validated. Up to now, these constraints have not been implemented in the German production system.

Consistency checks between Volume 1 and Volume 3

As between Volume 1 and Volume 2 there is also a need for consistency checks across Volume 1 and Volume 3. The following constraints should be applied:

Constraint C4: Codes in the Alphabetical Index must be valid nodes in the hierarchies of 4CH and MOR.

Implementation: Similar to T2

Constraint C5: If a code in the Tabular List is split into two or more "new" codes, all index entries referencing the "old" code must be checked.

Implementation: Similar to T5

Constraint C6: All changes made to inclusion and exclusion notes in the Tabular List must be checked against the Alphabetical Index.

Example: If texts of the inclusion or exclusion notes are updated in the Tabular List, the corresponding entries in the Alphabetical Index must be checked.

Implementation: The best way to ensure consistency for these changes, is a production system which lists all relevant entries from the index in a separate window when editing in the Tabular List and vice versa. The SGML based system cannot yet provide such an environment. In a database system such views should be created more easily.

Conclusion:

This paper presents a set of constraints to ensure consistency of ICD-10 during update procedures. These constraints are probably not yet complete and will have to be maintained over the years following problems which become evident during the update process. However, already implementation of some of these constraint will help to maintain consistency of ICD-10 over the years of the update process.

References

1. Schopen M: Electronic Publishing of the German Language Edition of ICD-10. WHO/GPE/ICD/C/98.24
2. Schopen M: Electronic Publishing of the Alphabetical Index to ICD-10. WHO/GPE/ICD/C/99.33

Dr. med. Michael Schopen
Deutsches Institut für Medizinische
Dokumentation und Information DIMDI
Waisenhausgasse 36 – 38 a
D-50676 K ö l n

Telephone: +49 221 4724-325
Telefax: +49 221 4724-444
e-mail: schopen@dimdi.de