WHO-FIC Classifications and Terminology Mapping

Principles and Best Practice
Acknowledgements

This document has been developed with support by the WHO-FIC Network, and in particular, Ann-Helene Almborg¹, Sharon Baker², Hazel Brear³, Can Celik⁴, Christopher G. Chute⁵, Vincenzo Della Mea⁶, Kathy Giannangelo⁷, Anupama Ginige⁸, Nick Hardiker⁹, Lara Harmans¹⁰, Robert Jakob⁴, Cassandra Linton², Jane Millar¹¹, Donna Morgan¹¹, Samson W. Tu¹², Coen van Gool¹⁰, Stefanie Weber¹³

¹National Board of Health and Welfare, Sweden, Nordic WHO-FIC Collaborating Centre; ²Canadian Institute for Health Information, Canadian WHO-FIC Collaborating Centre; ³NHS Digital, United Kingdom, UK WHO-FIC Collaborating Centre; ⁴Classifications and Terminologies, World Health Organization; ⁵Schools of Medicine, Public Health, and Nursing Johns Hopkins University, Baltimore, USA, WHO Collaborating Centre for Classifications, Terminologies and Standards; ⁶Dept. of Mathematics, Computer Science and Physics, University of Udine, Italy, Italian WHO-FIC Collaborating Centre; ⁷KG Consulting, USA (at time of formulation with SNOMED international); ⁸School of Computer, Data, and Mathematical Sciences Western Sydney University, Australia, Australian WHO-FIC Collaborating Centre; ⁹School of Human & Health Sciences, University of Huddersfield, UK, International Council of Nurses; ¹⁰National Institute for Public Health and the Environment, The Netherlands, Netherlands WHO-FIC Collaborating Centre; ¹¹SNOMED International; ¹²Center for Biomedical Informatics Research, Stanford University, USA, WHO Collaborating Centre for Classifications, Terminologies and Standards; ¹³Federal Institute for Drugs and Medical Devices, Germany, German WHO-FIC Collaborating Centre
**Table of Contents**

Executive Summary ............................................................................................................................................ 1

Principles and Best Practice .......................................................................................................................... 3

1. Introduction ............................................................................................................................................... 7

2. Purpose of the Paper ............................................................................................................................... 8

3. Definitions and Scope of the Paper ......................................................................................................... 10
   3.1 Definition of Terms ............................................................................................................................ 10
   3.2 Scope of the Paper ............................................................................................................................ 13

4. Background ............................................................................................................................................. 15
   4.1 Pre-Coordination and Post-Coordination ......................................................................................... 15
   4.2 Cardinality Ambiguities .................................................................................................................... 17
   4.3 ICD-11 ............................................................................................................................................... 18

5. Mapping Process ..................................................................................................................................... 20
   5.1 Overall Process .................................................................................................................................. 20
   5.2 Getting Started .................................................................................................................................. 22
   5.3 Selection of Mapping Techniques ..................................................................................................... 23
   5.4 Creation of Maps ............................................................................................................................... 23
   5.5 Documentation of Maps .................................................................................................................... 23
   5.6 Quality Assurance and Usage Validation of the Map ..................................................................... 23
   5.7 Dissemination of Maps ..................................................................................................................... 25
11. Appendices .............................................................................................................................................. 61

Appendix 1: ICD-10/ICD-11 Mapping Tools and Its Output Tabulation.......................................................... 61

Appendix 2: International Classification of Disease (ICD-10) to International Classification of Disease Australia Modification (ICD-10-AM) .............................................................................................................. 68

Appendix 3: SNOMED International, mapping SNOMED CT to ICD-10 .......................................................... 75

Appendix 4: NHS Digital (UK) SNOMED CT to ICD-10 Maps ......................................................................... 90

Appendix 5: Australian Classification of Health Interventions (ACHI) to International Classification of Health Interventions (ICHI) .............................................................................................................................................. 105

Appendix 6: International Classification of Diseases, Ninth Revision, Clinical Modification Volume 3 (ICD-9-CM Vol 3) to International Classification of Health Interventions (ICHI) .............................................................................................................. 110

Appendix 7: National Intervention Classification to the International Classification of Health Interventions .............................................................................................................................................. 118

Appendix 8: Swedish Classification of Health Care Interventions to the International Classification of Health Interventions .............................................................................................................................................. 127

Appendix 9: International Classification for Nursing Practice to the International Classification of Health Interventions .............................................................................................................................................. 134
Executive Summary

With the increasing adoption of electronic health records, the amount of coded data being collected increased exponentially. Today these data are encoded in a multitude of health care terminologies for different purposes. Terminology mapping is the method to transform such data so that they can be used for additional purposes, aggregated with data from other sources, or exchanged between systems. The World Health Organization Family of International Classifications (WHO-FIC) classifications, being the world standards for reporting diseases, health conditions, functioning status and (in the future) health interventions, play a central role in the storage, use, exchange, and analysis of data. To understand how these classifications and their terminological components can work with other health care terminologies and to translate data to and from these terminologies, this paper seeks to clarify the following:

- what terminology mapping is,
- how it is constructed, quality assured, validated and used,
- for what purpose it can be employed, and
- what limitations it may impose on the use of mapped data.

It recommends a collection of best practices and principles for the construction and use of terminology mappings.

The paper mostly follows International Organization for Standardization (ISO) standards to define terminology and terminology mapping and defines classification and clinical terminology as subclasses of terminology. The paper discusses the relation of these terms to terminology resource and ontology. For specifying the cardinality of mapping, the paper emphasizes the importance of clearly defining the source and target entities being mapped and the relationship represented by the mapping, especially in a world of post-coordinated terminologies, where a concept may be composed of multiple other concepts. The count of the target entities in a map may depend on whether a composed concept is considered as a single entity or as multiple entities. The paper also explains WHO's new architecture for developing and maintaining ICD-11, ICD and ICHI, which includes a knowledge base containing all entities relevant to ICD and a method to generate different variants of ICD from the knowledge base that requires a different way of thinking about mappings to ICD.
The paper draws on the co-authors' knowledge and experience and a literature search to describe the process of and techniques for developing terminology maps. The process description includes suggestions for best practices and emphasizes the iterative nature of map development and the importance of quality control and usage validation. The paper describes the steps involved in manual mapping and some of the technologies, such as natural language processing and machine learning, that have been applied to automate the generation of possible maps between terminologies.

Based on their prior work, the co-authors documented twelve case studies of terminology mappings. The case studies include two that involve mappings of different ICD versions, two that map Orpha-codes to ICD, two that map SNOMED CT to ICD, four that map national intervention classifications to ICHI, one that maps a nursing terminology to ICHI, and one that maps SNOMED CT to ICF. Nine of these case studies are available as appendices of this paper. Another one is extensively documented in a published paper.

The emergence of logic-based terminologies and ontologies opens up the possibility of developing and applying terminology maps using automated reasoning. The development of terminology maps should also be seen as an opportunity to identify gaps and possible errors in the source and target terminologies. At the same time, the differences in the intended purpose (use case), structure, and constraints of different terminologies imply that maps between them will often be approximate and dependent on the subjective judgement of map developers and users. Such limitations of terminology maps also mean that transformation of data based on these maps may involve information loss. Whether such information loss is desirable or can be mitigated depends on the intended usage of the data. Using terminology mapping to achieve data interoperability is also fraught with problems and wastes enormous resources.

The summary of the principles and best practices described in the paper is listed here as a set of bullet points.
Principles and Best Practice

This paper identifies principles and best practices over the lifecycle of a map. These are summarised here. Detailed descriptions of the context, procedures and explanations can be found in the following sections of this document.

1. Establish use case(s) before developing the map
2. Clearly define the purpose, scope, and directionality of the map.
3. Maps should be unidirectional and single purposed. Separate maps should be maintained for bidirectional maps (to support both a forward and a backward map table). Such unidirectional maps can be handy to support data continuity for epidemiological and longitudinal studies. Maps should not be reversed.
4. Develop clear and transparent documentation that is freely available to all and describes the purpose, scope, limitations, and methodology of the map.
5. Ideally, the producers of both terminologies in any map participate in the mapping effort to ensure that the result accurately reflects their terminologies' meaning and usage. At a minimum, both terminology producers should define the primary purpose and parameters of the mapping task, review and verify the map, develop the plan for testing and validation, and devise a cost-effective strategy for building, maintaining, and enhancing the map over time.
6. Map developers should agree on team members' competencies, knowledge, and skills at the project's onset. Ideally, users of the map also participate in its design and testing to ensure that it fits its intended purpose.
7. Quality Assurance (QA) and Usage Validation: QA and usage validation is ensuring the reproducibility, traceability, usability and comparability of the maps. Establish the QA and usage validation protocols at the beginning of the project and apply them throughout the mapping process. Factors that may be involved in quality assurance include quality-assurance rules, testing (test protocols, pilot testing), and quality metrics (such as computational metrics or precisely defined cardinality, equivalence, and conditionality). Usage validation of maps is an independent process involving users of the maps (not developer of the maps) in order to determine whether the maps are fit for purpose (e.g. do the end users reach to the correct code in the target terminology when using manual and automated maps etc.). Usage validation is essential to ensure the integrity of the information from source data to the final coding. Key principles for usage validation of maps include:
a. use of the ground truth of the original source data\(^1\) (e.g. diagnosis as written in the medical record) as the reference point;

b. compare the original source data with the end results of the following two processes

i. Coding of original source data with a source terminology – map code(s) of source terminology to code(s) of target terminology

ii. Coding of original source data with target terminology

c. statistically significant sample size that is representative of the target terminology and its prototypical use case settings.

d. Usage validation of automated maps should always include human (i.e. manual) validation

Clear documentation of the QA process and validation procedures is essential in this step in the mapping process.

If conducting a pilot test is feasible, it will improve the QA/validation process. Mapping is an iterative process that will improve overtime as it is used in real settings.

8. Dissemination: Upon publication and release, include information about release mechanisms, release cycle, versioning, source/target information, licence agreement requirements, and a feedback mechanism for users. Dissemination of maps should also include documentation, as stated above, describing the purpose, scope, limitations, and methodology used to create the maps.

9. Maintenance: establish an ongoing maintenance mechanism, release cycle, types and drivers of changes, and versioning of maps. The maintenance phase should include an outline of the overall lifecycle plan for the map, open transparent resolution mechanism for mapping problems, continuous improvement process, and decision process around when an update is required. Whenever maps are updated, the cycle of QA and validation must be repeated.

10. When conducting mapping manually, it is recommended to provide map specialists with the necessary tools and documentation to drive consistency when building the map. These include such items as the tooling environment (workflow details and resources related to both source and target schemes); source and target browsers, if available; technical specifications (use case, scope, definitions);

\(^{1}\) Not all mappings have “source data” e.g. local lab terminology to LOINC.
editorial mapping principles or rules to ensure consistency of the maps, particularly where human judgement is required; and implementation guidance. Additionally, it is best practice to provide an environment that supports dual independent authoring of maps as this is thought to reduce bias between human map specialists. Developing a consensus management process to aid in resolving discrepancies and complex issues is also beneficial.

11. In computational mapping, it is advisable to include resources to ensure consistency when building a map using a computational approach, including a description of the tooling environment, when human intervention would occur, documentation (e.g. the rules used in computerized algorithms), and implementation guidance. It is also advisable to always compute the accuracy and error rate of the maps. It is also essential to manually verify and validate the computer-generated mapping lists. Such manual checking is necessary for the quality assurance process, as maps generated automatically will almost always contain errors. Such manually verified maps can also help train the machine-learning model when maps for different sections of terminologies are being generated sequentially.

12. Cardinality is a metric in mapping that must be clearly defined regarding what is being linked between source and target and how the cardinalities are counted. For example, SNOMED CT codes for functional impairments are semantically different from ICF codes. A $1:1$ map between the two does not mean semantic equivalence. In terms of counting, what SNOMED International considers to be a $1:1$ map includes what others may consider being a $1:many$ map.

13. Level of equivalence, such as broader, narrower, or overlap, should be specified.

14. Maps must be machine-readable to optimize their utility.

15. ICD-11: When creating maps using ICD-11 and other WHO systems, mapping into the Foundation Component comes first, then maps to MMS could be created through linearization aggregation.

---

2 The detailed primary coding with URI or code combinations must be retained at the source to avoid data losses through linearization-maps
1. Introduction

Health care has always been a practice that depends on specialized vocabularies to name entities, such as symptoms, diagnoses, anatomical locations, procedures, and medications, for communicating and storing information. To transform information into data that can be analyzed, exchanged, aggregated, and processed by machines, it needs to be encoded in controlled terminologies. Over the years, numerous controlled terminologies have been developed to satisfy various requirements. These terminologies differ in their intended usage, coverage, structure, and governance. The interoperability of these terminologies has been a pressing issue for users of medical data and developers of health care applications and terminologies [1, 2]. For example, data coded in one terminology may need to be merged with data coded in another terminology; data collected for one purpose (e.g., clinical care and documentation) in one terminology may need to be mapped to another terminology that is more appropriate for another purpose (e.g., statistical aggregation, reimbursement, or clinical research).

With the increasing adoption of electronic health records [3, 4] and advances in information technology, the amount of coded data is increasing exponentially, either directly through structured data entry and automated reporting or indirectly with the use of natural language processing (NLP) techniques to extract machine-processable information from narrative texts [5-7]. Recently, with the explosion of genomic and other -omic data, the possibility of deriving clinical insights from biological data calls for the linkage of biological and clinical concepts [8] [9]. With these drivers for increased data integration and reuse, terminology mapping becomes even more urgent in the current healthcare landscape.

Another reason for the need for terminology mapping is the dynamic nature of health care. As scientific understanding of disease processes improves and new interventions become available, terminologies need to evolve. A prime example is the International Classification of Diseases (ICD) maintained by the World Health Organization (WHO). With the release of ICD-11 in June 2018 for member states and other stakeholders to use in preparation for implementation and translation, how ICD-11 relates to earlier versions of ICD and other terminologies becomes an important question.

The nature of terminologies is also changing. ICD-11 is being introduced in an era that is completely different from the era when ICD-10 was released. The “terminology
playing field” when ICD-10 was introduced looked nothing like it does now: not only are there more players now, but the players can also do more. Modern health care terminologies are no longer enumerations of hierarchically organized codes. Instead, terminologies are being coupled with ontologies that have formal semantics. The concept of post-coordination allows specifications of new concepts from combinations of existing concepts. Mapping as a simple correspondence between one set of terms with another no longer suffices. So we need to check how the rules of the game (mapping processes) need to evolve in the more complex world.

This paper focuses on terminology mapping to or from the classifications in the WHO Family of International Classifications (WHO-FIC) because of their central role in the world of interoperable data and reuse of data. ICD, ICF and ICHI are the most important and widely implemented standard terminology mandated by international regulations that serve as “the foundation for identifying health trends and statistics globally, and the international standard for reporting diseases and health conditions” [10]. The International Classification of Functioning, Disability and Health, more commonly known as ICF, is a foundational framework for measuring health and disability at individual and population levels. The International Classification of Health Interventions (ICHI) provides a standard for the international comparison of statistics on health care interventions. To understand how these classifications can work with other health care terminologies and how data can be translated to and from these terminologies, we need to have a clear understanding of what terminology mapping is, how it is constructed and used, for what purpose it can be employed, and what limitations it may impose on the use of mapped data.

2. Purpose of the Paper

This paper’s main intended audiences are the organizations and classification specialists who develop and maintain maps between health care terminologies and WHO-FIC classifications. In addition, government entities and statisticians attempting to analyze and compare coded data across various health care classifications and terminologies may find the case studies and the discussion in relation to cardinality useful. Furthermore, the paper may be of interest to researchers who are exploring different approaches and techniques to carry out mappings or improve the process of developing maps and the existing maps.
The paper summarises state of the art in terminology mapping as it exists now and explores how ICD-11, with its variant classifications generated from a common knowledge base, introduces new possibilities in developing terminology maps. The document provides an overview of the map development process and identifies the steps and stages associated with creating maps between two terminologies. This process description should assist map development managers in planning the mapping project and allocating resources as required. The description of various mapping techniques used by map developers worldwide should help mapping teams pick a suitable technique based on in-house expertise and resources for their mapping task. Most importantly, this paper establishes a shared vocabulary for describing health care terminology mapping, which should assist discussions and collaborations between mapping teams. The collection of case studies provided here summarizes current practices around the world to a wider audience. These case studies have aided in establishing the principles and best practices reported in this paper. These principles and best practices, if followed, should help teams who want to create new maps avoid pitfalls and reduce trial and error. In addition, the cautions and limitations described in the paper provide helpful information for statisticians and researchers analysing coded data and for classification mapping teams to consider when carrying out their own mapping exercises.

One new variable in terminology mapping is how WHO generates multiple variants of ICD-11 from entities in a knowledge base (called the Foundation Component) that include all concepts and relations relevant to ICD. The ICD-11 Mortality and Morbidity Statistics (MMS) tabulation, for example, is programmatically generated from the Foundation Component. This opens the possibility that mappings to ICD-11 may use Foundation Component entities as targets and programmatically derive maps to different variants of ICD-11 from the same set of mapping relations. While it is too early to give a ready-made recipe for performing such mappings, the case study of the mapping between ICD-10 and ICD-11 MMS described in this paper is an example of how mapping rules can be defined for entities in the Foundation Component. The experimental work to develop a common ontology shared by SNOMED CT and ICD [11-13] also points to a future where semantically sound ontological principles form the basis of terminology mappings.
Finally, the list of best practices extracted from the literature and the case studies of mappings presented here is another contribution of this paper. These best practices are grounded in the case study authors’ experiences and supplement those top-down principles enumerated in works such as ISO/TR 12300:2014 [14]. They provide empirically derived guidance to practitioners of terminology mappings.

3. Definitions and Scope of the Paper

3.1 Definition of Terms

ISO 1087-1:2000 [15] defines terminology as a set of designations (i.e., representations of concepts by signs which denotes them) belonging to one particular language. In this paper's context, the special language is the language for describing entities and relationships in the health care domain. Following the "Family of International Classifications" paper [16], we further distinguish a classification (‘an exhaustive set of mutually exclusive categories to aggregate data at a pre-prescribed level of specialization for a specific purpose’) and a clinical terminology (‘terminology required directly or indirectly to describe health conditions and health care activities’) as special types of terminologies. In the context of this paper, a classification consists of mono-hierarchies that have residual classes at each level to ensure that, in a statistical aggregation, entities in a hierarchy are not counted double and that every phenomenon in the domain can be assigned to a category in the classification. To support clinical activities documentation at the level of detail required for health care, concepts in a clinical terminology like ICD or ICHI foundation, SNOMED CT or ICF foundation tend to be more granular than those in a pure classification.

The designations in a terminology may be non-semantic concept identifiers (i.e., codes) that satisfy Cimino’s desiderata for controlled vocabularies [17]. To be human-understandable, concepts need to have terms, language-specific phrases that convey the concepts' intended meaning. The description of a concept in a terminology may include additional features, such as definitions, synonyms, and coding advice.

ISO/TR 12300:2014 [14] defines terminology mapping as the "process of defining a relationship between concepts in one coding system to concepts in another coding system, following a documented rationale, for a given purpose." A map is either an individual map (i.e., "index from one term to another") or a map set (i.e., a “group of
individual maps used to convert a range of entries from source to target code”). A map necessarily has directionality: a map source is related to one or more map targets.

We distinguish between (1) mapping between similar semantic domains, such as mapping of SNOMED CT disease concepts to ICD disease concepts, where it is usually meaningful to characterize a source concept as semantically equivalent to, broader than, narrower than, or overlap with the target concepts; and (2) mapping between different semantic domains, where a source concept is linked to target concepts that are different from the source concept. An example of a mapping between different semantic domains is a mapping between a pharmaceutical product terminology and an ontology of chemicals, where a product like Fluoxetine 4 MG/ML Oral Solution may be linked to the Fluoxetine substance through a has_ingredient link. The literature on terminology mapping often implicitly assumes that the mapping under discussion is of the first kind (i.e., mapping between similar semantic domains). Sometimes the distinction may be subtle. An example of mapping between different semantic domains is a mapping that links SNOMED CT’s Pain (SCTID: 22253000) to ICF’s pain concept [18]. However, SNOMED CT’s Pain (SCTID: 22253000) is a kind of finding, which represents the result of a clinical observation, assessment or judgment [19]. The concept of pain in ICF, on the other hand, is a kind of body function that can be used to denote a normal or abnormal state only if a qualifier is added. To avoid this kind of ambiguity, terminology mappings should specify the specific relationship between source concepts and target concepts in a map. For example, a SNOMED CT disease (disorder) may be equivalent to, broader than, narrower than, or overlap with an ICD disease concept. In contrast, a SNOMED CT finding may involve an ICF body function.

Even when the source and target terminologies are in the same semantic domain, the mapping relationship may not be of the is-a kind where the notions of generalization and specialization make sense. For example, the ICD-9 CM Volume 3 code 79.31 Open reduction of fracture with internal fixation – humerus represents a complex intervention that involves two actions: open reduction and internal fixation. In ICHI Beta-2 2019, we find two separate interventions: MEB.LD.AA Open reduction of humerus and MEB.DN.AA Insertion of device into bone of humerus where the ‘code also’ instruction for MEB.LD.AA suggests that an additional code may be assigned if an internal fixation is also performed. To say that ICD-9-CM code 79.31 has a one-to-many relationship to ICHI’s MEB.LD.AA and MEB.DN.AA obscures the fact that the procedures open reduction of humerus and insertion of device into bone of
humerus are not narrower than the open reduction of fracture with internal fixation procedure but are component parts of it.

A map may have additional attributes, such as conditionality (e.g., SNOMED CT’s Infertile (finding) (SCT 8619003) can be mapped to ICD-10’s N97.9, Female infertility, unspecified or N46, Male infertility, depending on whether the source context is female or male), and confidence level (e.g., when the map is created using an automated method and similarity scores).

In addition to issues related to the semantic match of source and target, a map may be complex due to the characteristics and usage rules of the source or target terminology. Terminology mapping is not just the alignment of concepts based on their meaning. There must be an expert understanding of the business rules associated with both source and target. For example, clinical records for an encounter must be mapped to a classification in such a way as to include crucial statistical characteristics such as 'Other specified', 'Unspecified' and 'Not otherwise classified.' These types of entities are not appropriate for a clinical terminology and should not usually appear in a patient medical record. However, they are meaningful in a statistical record. Consequently, the application of business rules may change the code assignment applied using semantic matching alone. Mappings to classifications, most importantly, should ensure that all target codes are mutually exclusive for counting purposes.

Appendix 4 (Mapping from SNOMED CT to ICD-10: UK Edition) shows another example where the SNOMED CT concept 10698009 | Herpes zoster iridocyclitis (disorder) may be linked to a choice of classification maps, depending on whether herpes zoster or iridocyclitis should occupy the position of primary code in ICD-10’s dagger-and-asterisk combination. This map’s design allows the coder to apply the primary morbidity focus of care business rule and assign the correct code or codes dependent on information that is only available at the point of coding.

Terminology mapping, the process of translating one set of codes to another, should be distinguished from recoding, where a coder has access to the primary data record that had been coded in the first terminology and chooses appropriate codes using a second terminology. The resulting codes may or may not be the same as codes derived from mapping the first terminology to the second terminology. Coding may depend on patient contexts that are no longer available at the time of terminology mapping.
In the literature on terminology mapping, two additional terms warrant clarification: (1) a **terminology resource** that includes not only terminologies as defined above, but also lexical databases such as WordNet [20] and multi-terminology metathesaurus such as the Unified Medical Language System (UMLS) [21]; (2) an **ontology** which we define as a logic-based representational artefact that formalizes the entities and relationships in a domain of discourse [22, 23]. We adopt a relatively narrow definition of ontology, requiring that it models statements about entities in the domain that are necessarily true without exception. Large medical terminologies often contain an ontology but also include elements that are not ontological in our sense. For example, the National Drug File-Reference Terminology (NDF-RT) [24], even though it is represented in the Web Ontology Language (OWL), contains a large amount of non-ontological facts, such as the diseases that a drug may treat or may prevent. SNOMED CT, based on a variant of description logic [25], similarly includes concepts such as *situation with explicit context* that mix characteristics of information models with characteristics of ontologies [26]. Additionally, SNOMED CT concepts may include epistemological concerns, such as degree of certainty and modality (e.g., 12601000132105 | Probable severe acute respiratory syndrome (situation) and 102874004 | possible pregnancy), that have no place in a strict ontology [27].

### 3.2 Scope of the Paper

The focus of the paper is on mappings that involve classifications within the WHO-FIC, defined in the Denny et al. [16] paper as consisting of three reference classifications (ICD, ICF, and ICHI), derived classifications such as the International Classification of Diseases for Oncology (ICD-O-3) [28], and related classifications such as the International Classification of Nursing Practice (ICNP) [29].

**Use cases**

In this paper, we define use cases of terminology mapping primarily in terms of the goals or purposes that a user or a software system can achieve using specific terminology mappings. A literature search revealed a myriad of purposes that mappings between specific terminologies or terminology resources have been used for. Some of the usages that may be relevant to WHO-FIC classifications include:

- **Transformation of data for different usage**: e.g., Campbell et al.[30] and Giannangelo et al. [31] map SNOMED CT to ICD-10 to translate data collected for
clinical documentation to a classification appropriate for epidemiology; terminology mappings enabled Declerck et al. [32] and Fung et al. [33] to use clinical data for pharmacovigilance research and billing respectively.

- **Use class hierarchies of one terminology to aggregate data coded in a second terminology**: e.g., Pathak et al. [34-36] investigated whether it’s possible to use NDF-RT axes to aggregate and analyze generic and branded drug products coded in RxNorm.

- **Bridging data coded using different versions of a terminology**: e.g., Rakhsi-Raiez et al. [37] mapped versions of intensive-care classifications.

- **Identifying gaps in, and potentially enriching terminologies with, terms from other terminologies**: e.g., He et al. [38] used UMLS to find potential additions to the National Cancer Institute Thesaurus.

- **Migration of legacy data**: e.g., Fung et al. [39] developed automated methods to translate clinical phenotypes defined in ICD-9-CM to ICD-10-CM.

- **Integration of data sets that use different terminologies**: e.g., Ficheur et al. [40] developed an automated method that uses the distributional properties of laboratory parameters to map parameters from different hospitals; Peters et al. [41] used RxNorm to integrate data coded using local drug terminologies.

- **Construction of a terminology subset by mapping an existing specialized terminology into a standard terminology**: e.g., Bakhshi-Raiez et al. [42] mapped a terminology in the intensive care domain to SNOMED CT to derive a subset of SNOMED CT for use in intensive care.

Mapping clinical terminologies to classifications for aggregation and data reuse are classic use cases for terminology mappings involving WHO-FIC classifications. Similarly, bridging data coded using different versions of WHO-FIC classifications (e.g., national adaptations to the international versions and between versions of ICD) is a requirement for using WHO-FIC classifications. Mapping of national intervention terminologies to the emerging ICHI to identify potential gaps in ICHI is part of ICHI testing. This paper will focus on these use cases as the primary use cases. Others will be considered secondary use cases.

Out of scope for this paper are mappings where both the source and target terminologies are not members of WHO-FIC classifications. Mappings between versions of national adaptations of WHO-FIC classifications are not in scope.
4. Background

Before we delve into the process, techniques, case studies, and the opportunities and limitations of terminology mappings, we first give some background on the concept of post-coordination and the innovations of ICD-11 and explain the issue of mapping cardinality for which we recommend some best practices that are not in the current literature.

Historically, terminology mapping involves selecting codes from a fixed collection of codes in the target terminology that best match a given source code according to a set of criteria. In this context, descriptions of maps can make assumptions, such as a one-to-many map implying that the union of the target codes best approximate the meaning of the source code. However, when terminologies allow the use of multiple codes to specify a concept, a process called post-coordination, the notion of mapping cardinality becomes ambiguous and is used inconsistently in different descriptions of terminology mapping.

As described earlier, ICD-11 introduces a new factor that affects its role as a mapping target. Instead of having a primary classification and derived classifications that are maintained separately, ICD-11 introduces a Foundation Component that encompasses concepts relevant to all variants of ICD and derives alternative linearizations for different use cases.

4.1 Pre-Coordination and Post-Coordination

Pre-coordinating concepts in a terminology is to have a fixed enumeration of all possible concepts in the terminology. For example, infectious diseases may be described by the causal agent (bacterial, virus, etc.), anatomical location, and temporal course. Pre-coordinating the hierarchy of infectious diseases means enumerating all sensible combinations of the causal agent, anatomical location, and temporal course and making them individual concepts represented in terminology. The terminology would be populated by terms such as acute-viral-lung disease, acute-bacterial-intestine disease, etc.

Post-coordinating a terminology is to have a method to combine multiple concepts from the terminology to represent concepts that do not have fixed identifiers in the terminology. The method to form new concepts may use logical operations like a
conjunction, disjunction, or negation, relations such as associated with, or it may involve multiple axes of descriptors (aka extension codes) that can be chosen to modify an existing concept (aka stem code). Post-coordinating a terminology about infectious diseases may involve creating explicit hierarchies of infectious organisms, anatomical locations, and possible temporal courses. A user finds the equivalent of acute-viral-lung disease by selecting appropriate descriptors in the temporal course value set (e.g., acute), infectious organism hierarchy (e.g., virus) and anatomical location hierarchy (e.g., lung).

ICD-10, on occasion, does allow the use of multiple codes to specify related information. First, there is the dagger-and-asterisk system where a dagger code representing an underlying generalized disease (e.g., A18.1† Tuberculosis of genitourinary system) may be combined with an asterisk code that represents a manifestation in a particular organ (e.g., N33.0* bladder) or site that is a clinical problem in its own right. ICD-10 allows additional dual coding in other circumscribed situations. For example, the block of categories B95-B98 specifies bacterial, viral, and other infectious agents that can be used to identify the infecting organisms in local infections that are classifiable to the ‘body systems’ chapters. For neoplasms with functional activity, the codes from the endocrine, nutritional and metabolic diseases chapters may be added to the code from the neoplasm chapter to indicate the type of functional activity. For example, C74 catecholamine-producing malignant phaeochromocytoma of adrenal gland can be augmented with the code E27.5 Adrenomedullary hyperfunction. Volume 2 of the ICD-10 enumerates these and other opportunities for multiple coding.

In place of ICD-10’s ad hoc mechanisms for post-coordination, ICD-11 introduces explicit post-coordination in the form of associations between ICD codes and the use of stem codes and extension codes (see the following section on ICD-11).

Coding a subject’s functional status using the International Classification of Functioning, Disability, and Health (ICF) can be seen as a kind of post-coordination as well. ICF organizes information in a collection of orthogonal components: body functions (b codes), body structures (s codes), activities and participations (d codes), and environmental factors (e codes). Each component has several component-specific qualifiers (e.g., extent, nature, and location of an impairment for the body structure component). Coding ‘Able to walk for more than 1 mile with hand-held assistive device in one hand,’ for example, requires the following combination of codes:
4.2 Cardinality Ambiguities

ISO/TR 12300:2014(E) [14] defines the cardinality of a map as an important characteristic of a map. The cardinality of a map (one-to-one, one-to-many, many-to-one, and many-to-many), however, has a very weak semantic definition, being nothing more than the numbers of source entities and target entities that are linked in the map. The best use of map cardinality involves maps where the source and target terminologies are pre-coordinated, the "link" is limited to the degree of semantic equivalence, and the "many" concepts in source and target terminologies are implicitly joined as a disjunction. For example, in the mapping of ICD-9-CM Vol 3 to ICHI (Appendix 6), the many-to-one mapping of Laparoscopic appendectomy (47.01) and Laparoscopic incidental appendectomy (47.11) to Laparoscopic appendectomy implicitly assumes that the union of the source concepts approximates the target concept more closely than either of the source concepts alone. Thus, a many-to-one map usually means that each of the source concepts is narrower in meaning than the target concept, and a one-to-many map usually means that the source concept is broader in meaning than each of the target concepts. However, whether a source concept is broader or narrower than a target concept is not guaranteed, as the union of the "many" concepts may be broader than the "one" concept. In that situation, one of the "many" concepts may just overlap in meaning with the "one" concept. Such ambiguity is also present in a one-to-one map, where the map may mean the source concept is equivalent to, broader than, narrower than, or overlaps with the target concept. As discussed earlier, the ambiguity in the specification of the map, cardinality is worsened when the source and target terminologies have different semantic domains.

In a world where both the source and target concepts may be post-coordinated, the notion of "one" and "many" become even more ambiguous, and a clear specification of the semantics of the "link" becomes more important. Does the post-coordinated ICD-11
expression CA40.1 (Viral pneumonia)\XN8W9 (Pneumovirus) represent a single concept or two? If the post-coordination logic allows disjunction, the meaning of “many” in map cardinality requires clarification. If a terminology, such as ICF, is multiaxial, mapping to or from such a terminology requires the map to clearly label the semantics of the “link” involved in the map. In case studies presented in this paper, we will treat expressions that involve a stem code and one or more extension codes as single concepts and avoid expressions that involve disjunctions of concepts.

Given the possible ambiguities in the definition of map cardinalities, it may be better to eschew their use, and define mappings purely in logical terms, as is done in the ICD-10/ICD-11 mapping (Appendix 1), where the relationship between source and target concepts are specified as equivalence, subset, or superset explicitly. When multiple concepts participate as source or target of the map, they are defined as a union of the concepts involved.

We recognize that in some use cases, the map cardinality expressed in numerals (e.g., 1:3 or 1:4) can be useful when the units of comparison and the semantics of the "link" between source and target terminologies are well defined. For example, for mapping an existing well-structured national classification to a WHO reference classification or for mapping between versions of WHO reference classifications, the codes and their usage rules are compatible, and the meaning of mapping cardinalities is clear. In this case, the purpose of the mapping is to convert to or use the WHO reference classification or for statistical follow-up of trends after a change of classifications.

4.3 ICD-11

The Foundation Component is a knowledge base that contains all the information related to ICD, including ICD-10, ICD-11, and (eventually) national modifications. It systematically represents the scientific, linguistic, and classification knowledge about diseases, disorders, injuries, and other health-related concepts within the scope of ICD. The Foundation Component is organized in a poly-hierarchy where the nodes are arranged by the principle of generalization and specialization. *Influenza due to identifiable influenza virus*, for example, should be classified not only as a disease of the respiratory system, but also as a kind of infection. In addition, the Foundation Component contains the specifications for programmatically generating linearizations, such as tabularizations for Mortality and Morbidity Statistics (MMS) and Primary Care - Low Resource Setting.
The linearizations differ importantly from the Foundation Component in their properties and applications. A linearization, such as the MMS Tabulation, is intended to function as statistical classification, which implies that the sibling categories at any level must be mutually exclusive and exhaustive. The mutual exclusivity constraint is obtained by requiring that linearizations must comprise mono-hierarchies, meaning that an ICD category must have one and only one parent to avoid statistics that are distorted by double counting. The exhaustivity constraint is achieved by adding residual categories, such as other specified or unspecified, at the terminal leaves of the mono-hierarchy to ensure that any encountered condition has a place to be counted. There are no residual categories in the Foundation Component; thus, it is technically not a superset of all the linearizations. However, any category other than residual categories must appear in the Foundation Component for it to be included in a linearization.

The roles of Foundation Component and linearizations in ICD-11 opens the possibility that terminology mappings that have ICD-11 Foundation Component as the target can have mappings to different variants of ICD-11 generated automatically. In fact, this is how the ICD-10 mappings to and from ICD-11 MMS Tabulation are done (see Appendix 1).

ICD-11 differs from previous revisions in another fundamental way. It has full support and, in some cases, requires post-coordination of "stem" codes with modifiers (also known as extension codes) or other stem codes to form composite concepts. For example, most ICD-11 disease and disorder codes can be modified to indicate specific anatomy, laterality, or severity. Diabetes can be "clustered" with renal failure in a post-coordinated expression, indicating that diabetes has resulted in renal failure as a complication. Furthermore, ICD-11 supports some limited sanctioning to disallow nonsensical expressions and post-coordinated expressions that are equivalent to existing pre-coordinated codes.

In the next two sections, we describe the process and techniques of developing terminology maps.
5. Mapping Process

5.1 Overall Process

The high-level process for creating and maintaining a map between two terminologies is depicted in Figure 1. We focus on seven steps in the process: (1) Getting Started, (2) Selection of Mapping Techniques, (3) Creation (or Revision) of Maps, (4) Documentation of Maps, (5) Quality Assurance and Usage Validation of Maps, (6) Dissemination of Maps, and (7) Maintenance. While this depiction recommends the sequential progression of these steps, the ordering of the steps could be flexible depending on the work that has been done before. The process of developing and maintaining a terminology map is inherently iterative [43], as shown by the backwards-pointing arrow in the diagram. Steps 2 to 5 can be performed at the beginning as a pilot study that validates the process and documentation and that informs the subsequent map development.

Irrespective of the maturity of the mapping (i.e. whether the maps currently exist or not), reassessing the purpose and scope (Step 1) of the maps and carrying out a validation task (Step 5) would help to improve the quality of the maps. In some cases, using a different mapping technique may assist in identifying inaccuracies and fixing these. For example, in the case of mapping ICD-10-AM to/from ICD-10, the maps were originally created manually, and re-mapping with a computerized technique has helped fix human-made errors (see Appendix 2).
To create and maintain high-quality terminology maps, quality assurance and usage validation protocols for the mapping process should be agreed upon at the beginning of a mapping project. They are crucial for enabling the creation of a reliable, consistent and reproducible link from one step of the mapping process to the other. It is good to bear in mind that different methods of mapping may have different quality assurance documentation needs and may require different approaches. ISO/TR 12300:2014 [14] is a good reference for factors that affect the quality and cost of a terminology mapping project.

Ensuring the quality of the mapping process requires the planning and documenting of all the steps and the choices that need to be made during the process of creating a map. It requires the definition and description of the scope, purpose, and stakeholders
of the mapping, the intellectual property arrangements when applicable, the source, target terminologies of the map, including their versions, the format of the map, and the plan for versioning, dissemination, maintenance and revision.

5.2 Getting Started

For the generation of any map, the first step is to identify the source and target schemes as well as the defined purpose and scope of the map.

5.2.1 Defining the purpose

The use case(s) for the map and the structure of the source and target terminologies, which will have their own purpose (separate from the map), will drive the requirements for map attributes, metadata, heuristics, advice and guidelines. The use case may demand the creation of a map for a one-time mapping or for a map that is regularly revised, released and maintained. There may be more than one use case applicable to a map, and the creators and end-users need to understand that there are risks involved with a multi-use map.

Documentation should contain scenarios that explain the intended users and intended use or uses of the map and that provide examples of where and how the map could be used. Common use case examples of maps (case studies) are summarized in Section 7, and more detailed descriptions are included as appendices to this paper.

5.2.2 Defining the scope

Part of getting started with the mapping process is to define what the scope of the map should be. Given the purpose of the map and the scopes of the respective source and target terminologies, subsets of the terminologies may be in-scope for the map; for instance, a map from a clinical terminology like SNOMED CT to a classification like ICD may declare only a subset of SNOMED CT in-scope whilst declaring the complete set of classification codes in-scope of a map. Any out-of-scope target maps should be clearly documented.
5.3 Selection of Mapping Techniques

Selection of the mapping technique will be dependent on the purpose and scope of the map being created, e.g. one-time use case. The common mapping techniques are described in Section 6 of the paper.

5.4 Creation of Maps

Resources for the creation of a map will be dependent on the mapping technique selected, and a mapping toolkit should be designed with the purpose, scope, source, target and selected technique in mind.

5.5 Documentation of Maps

Documentation should be clear, transparent and freely available to all. It should perfectly describe the purpose and scope of the map as well as provide insight into all processes which have had an impact on the assignment of target codes.

5.6 Quality Assurance and Usage Validation of the Map

The mapping product itself requires quality assurance and validation procedures to assure appropriate results for the intended use of the map. First, the mapping product can be evaluated in terms of an internal quality assurance process: the extent to which the intended mapping method is applied to the source and target terminologies. This answers the question of whether the chosen method of mapping was a fit in terms of source and target terminology characteristics. Emerging aspects in the specific editorial mapping principles and mapping workflow can result in different quality of maps in early stages of the mapping process compared to later stages. Second, the mapping product can be evaluated in an usage validation process: an independent but essential step to finalize maps, involving users of the maps (not developer of the maps) in order to determine whether the maps are fit for purpose (e.g. do the end users reach to the correct code in the target terminology when using manual and automated maps).

For usage validation of maps the following key principles apply
2. The ground truth is what the original source data says (e.g. diagnosis as written in the medical record) not what the source terminology coded record says.

3. Usage validation of maps should compare the original source data with the end results of the following two processes:
   a. Coding of original source data with a source terminology – map code(s) of source terminology to code(s) of target terminology
   b. Coding of original source data with target terminology
   In case the two process render different results (i.e. a mismatch between the end results is observed a discrepancy check and interview should be performed to determine the reasons for the mismatch.

• Usage validation of maps must be conducted with a statistically significant sample size that is representative of the target terminology and its prototypical use case settings. For example, when using ICD-11 as a target terminology the sample must include a diagnostic statement in the original source data which are representative in terms of ICD-11 chapters and coding situations (i.e. hospital inpatients, outpatients, primary care).

• The above listed principles are applicable for manual and automated maps. Usage validation of automated maps should always include human (i.e. manual) validation.

Examples of how mappings created or assisted by computational algorithms are evaluated against manually created reference maps are given in Section 6) and they are also found in the literature. For example, Reich et al. [44] converted ICD-9-CM diagnosis codes used in some existing databases to SNOMED CT and MedDRA and evaluated the impact of the mapping on the prevalence of outcomes and on the reliability of analytic methods designed to detect drug-outcome associations in test databases. Souvignet et al. [45] mapped MedDRA to SNOMED CT to create and evaluate a formal ontology by comparing the results of making queries with existing groupings. More recently, Hripcsak et al. [46] showed that extraction of eMERGE network phenotypes could be minimally affected by translating ICD-9-CM/ICD-10-CM encoded data into SNOMED CT encoding for phenotypes whose definitions did not involve multiple ICD-9-CM and ICD-10-CM codes being mapped to the same SNOMED CT codes.

Overall, there are several factors of quality assurance that come into play in the production phase of the mapping:
1. Quality assurance rules: these can include technical rules that may need to be developed as "must be fixed," whilst warnings or prompts may require confirmation of accuracy before allowing finalization of the map. A consensus management procedure is also an example of a validation rule in the case of simultaneous mapping by multiple mappers. Rules and warnings which are applied automatically and in real-time throughout the mapping process are desirable.

2. Testing and re-testing: A good description of the testing protocol(s) and iteration process(es) gives the user an idea of the quality of the production of the map and its progress; it also includes the phases of pilot testing and testing in the targeted use case environment and the description of situations when human intervention is necessary.

3. Quality metrics: can include computational metrics in case of algorithmic or another computer-assisted mapping, such as Precision, Recall, F-score and Accuracy \[47\]. In the case of manually constructed maps, the evaluation of a map’s quality depends on the degree to which mapping concepts such as cardinality, equivalence and conditionality are precisely defined.

5.7 Dissemination of Maps

Publication and release information for the maps should include details of versions of the source and target terminologies, distribution platform and mechanisms, frequency of release and license agreement requirements. The format of the map needs to be computer processable and clearly documented.

5.8 Maintenance and Revision of Maps

Maps can be used in different ways and with different frequencies; nevertheless, they broadly fall into two groups: one-time maps and maintained maps. To provide clear expectations to users with regard to the drivers for change, maintained maps should have a continuous improvement process by way of a maintenance and revision plan, indicating the lifecycle of the mapping \[14\].
Maintenance may involve correction of errors, resolution of mapping problems, revision of maps due to changes in the source and/or target terminologies, implementation of routine updates, or simply a review of a specific cohort of maps, which may or may not require amendment, re-testing and re-releasing.

The responsibility for the frequency of updates and type of changes should be clearly stated from the outset; most often, the terminology that is updated most often, either the source or the target, determines the pace of changes to the map. Other examples of drivers for maintenance include not only international and national directives and requirements but also regular reviews from users of the map or stakeholder groups.

6. Mapping Techniques

The literature on techniques for mapping terminologies is vast. Nikiema et al. [48] cite [49] as grouping these techniques into four methodological approaches: (i) manual approaches, (ii) morphosyntactic approaches, (iii) approaches based on semantic features of concepts (subsumption, roles, etc.), and the structure of terminology resource to be aligned and, finally, (iv) approaches using a third terminology resource as background knowledge. In this paper, we briefly survey the manual and computational approaches to creating maps.

6.1 Manual Mapping

There is a rich tradition of manual mapping involving WHO-FIC classifications, as described in the literature (e.g., [30, 31, 50]) and as exemplified by case studies summarized in and appended to this paper. Several authors of this paper have participated in numerous manual mapping exercises. In this section, they have distilled their knowledge and experiences to present some of the best practices in manually creating a terminology map.

For the purposes of this paper, we use "manual mapping" to describe mappings that are created by human map specialists using defined resources related to the source and target terminologies. Manual mapping involves the use of human knowledge and skill to build maps between concepts and/or terms in different terminological resources. Each map record is constructed individually. The process requires examination of every concept in scope within the source and target coding system. Informed judgements or decisions are made about the shared meaning of concepts [14].
6.1.1 Manual mapping ‘toolkit’

Once the purpose/purposes and scope of the map have been agreed upon, attention should turn to the resources that need to be provided to support and guide the manual mapping process, and that will increase the quality and integrity of the map. A mapping toolkit, such as the one described below, will aid interpretation of the source and target concepts and drive consistency when building the map.

6.1.1.1 Tooling environment

The toolkit should include a tooling environment that ideally supports the mapping workflow and contains the knowledge and best practice resources specific to both the source and target schemes, e.g. index, tabular list (e.g., ICD-10), reference guide, heuristics for creating maps, conventions, international and national editorial principles and relevant standards, where applicable. Map developers may wish to use a tooling environment that supports dual independent mapping as this is thought to reduce bias when selecting target codes.

6.1.1.2 Browsers

Source and target browsers, if available, are invaluable for understanding the meaning of the entities being linked and may often be available in the map tooling environment. Browser versions must correspond with the versions of source and/or target schemes being used for mapping.

6.1.1.3 Documentation

Documentation, such as a technical specification, editorial principles and implementation guidance, should identify the intended audience, for example, technical professionals and informatics specialists, and are crucial for supporting not only the map development but also for informing the use of the map to fulfil the purpose or purposes of the map. Documentation is also key in the quality assurance and validation processes. Each mapping project will define the types of documentation required. For example:

6.1.1.3.1 Technical specification

A Technical Guide provides the foundations for the construction of a map. It should identify the intended audience (e.g., technical professionals) and explain the use case, the scope of mapping, the definitions of map metadata, heuristics for target code
assignment, how to implement the map files as well as describing the release format of the map. The inclusion of accompanying examples would be desirable.

6.1.1.3.2 Editorial mapping principles

For a manual mapping project, editorial mapping principles are crucial. These principles should be designed to promote consistency, reproducibility and comparability. A principle will instruct the human mapping specialist during action in specific circumstances, for instance, in a judgemental assignment. Off-target code may lead to inconsistency between map specialists. Editorial mapping principles should be transparent and available to a user of the map, as confidence in a map will be increased when the thought processes and rationales behind target code assignments are available. The version of the editorial guide should correspond to the version of the map in which the guidance was used in its development.

6.1.1.3.3 Implementation/User guide

An implementation or user guide informs the end-user about the content and structure of the map and describes how the map should be used. It describes any editorial principles or other processes which have been applied in the construction of the map. Accompanying examples are particularly useful.

6.1.2 Consensus management

A defined consensus management process will support a manual mapping approach by aiding the resolution of discrepancies and complex issues. Additionally, emerging themes may be a driver for the development of editorial mapping principles. It is important to have an audit trail of all consensus discussions, research results and decisions taken around the assignment of targets.

6.1.3 Manual mapping steps

Recommended methodology for this approach is as follows:

3 Where source concepts that are being mapped to another terminology are specified uniquely and exactly in the target resources, a code assignment is unequivocal. All other code assignments are, to a greater or lesser extent, judgemental with varying degrees of ambiguity in the extent to which target concepts correspond to source concepts.
1. Evaluate source terminology/classification entity to understand the semantic domain.
2. Locate the best place for that semantic entity in the target terminology/classification.
3. Identify a default target code or codes ensuring application of the rules and conventions of the target scheme.
4. Consider other resources which may influence target code selection, such as guidelines, standards, definitions and documented requirements.
5. Apply approved editorial mapping principles.

6.1.4 Feedback mechanism

Providing a mechanism for users to feedback any issues is important. Any maps which have judgemental assignments would ideally be flagged as such, along with any mapping principles which have been clearly documented. This will enable the end-user to raise a query and will encourage stakeholders to positively participate in improvements, thus leading to greater map accuracy.

6.1.5 Competencies, knowledge and skills

Map developers should agree on the competencies, knowledge and skills required of the team members, such as map specialist, at the beginning of the mapping project. It would be difficult to list requirements here as they will vary, not only for each mapping project but also for each country and organization.

Nevertheless, target users of the map must participate in its design and testing to ensure that it is fit for its intended purpose [43].

Ideally, the producers of both terminologies in any map should participate in the mapping effort to ensure that the result accurately reflects the meaning and usage of their terminologies. At a minimum, both terminology producers must participate in defining the basic purpose and parameters of the mapping task, reviewing and verifying the map, developing the plan for testing and validation, and devising a cost-effective strategy for building, maintaining, and enhancing the map over time [43].
6.1.6 Evaluation of maps

In the manual mapping approach, the internal evaluation of mapping outputs is embedded within the mapping process described here and in the previous section on quality assurance and usage validation. Using editorial mapping principles and a mechanism for consensus and feedback management, ensure that the maps are evaluated during and after the mapping process. The usage validation of a map is discussed in Section 5.

6.2 Computational Mapping

We categorize computational techniques for creating not so much by the types of data or semantic structures involved in the mappings, as Saitwal et al. [49] had done, but by the computational techniques involved. First, because so much of the semantic information in the source and target terminologies is expressed in natural language strings and text, natural language processing (NLP) techniques are often employed in any computational approach for creating maps. Second, we highlight the use of numeric matching or scoring algorithms as a method to generate and evaluate potential individual maps between source and target terms. Third, we briefly describe the “rule-based” approach (non-numeric algorithms) for generating terminology maps. Finally, we describe the steps involved in the class of machine-learning techniques.

The limitation of using automated techniques is that the results are necessarily approximate. There will be false negatives (mappings can exist but not be made) and some false positives (mappings that are not correct). Thus human verification is an essential step in the computational approach to creating maps.

Just as a toolkit is needed in manual mapping, resources are needed to ensure consistency when building the map using the computational approach. This includes a description of the tooling environment and when human interventions in the mapping workflow occur. Documentation may include definitions of the rules used in the computerized algorithm as well as guidance for implementers and users. The role of the human reviewer should be defined, e.g., review of computer-generated maps to improve the algorithm or to train the model, along with the resources used, such as editorial mapping principles.
6.2.1 Natural language processing

NLP techniques are used to normalize and parse strings associated with the source and target concepts. The steps may include (1) Augmenting source and target terminologies with synonyms; (2) Tokenizing the characters in a term or its synonyms; (3) Normalizing a term or its synonyms by expanding acronyms, abbreviations, or shortened names, applying spelling corrections, removing stop words, and performing morphological analysis such as handling punctuation and other morphological variations; (4) Possibly parsing text and identify a part of speech (e.g., Elkin et al. [51] identified verbs as candidates for attributes of SNOMED CT concepts).

The application of NLP may require the use of other controlled terminologies (e.g. MeSH (Medical Subject Headings) [52] or UMLS (Unified Medical Language System) [53] and NLP systems dedicated for clinical text processing (e.g. Apache cTAKES - clinical Text Analysis and Knowledge Extraction System [54]).

6.2.2 Use of similarity metrics

A common approach to automate the generation of mapping candidates is to use similarity metrics. This approach first involves using NLP techniques to process terms and descriptions from the source and target terminologies, then algorithmically generates candidate pairs for individual maps, possibly filtering candidate pairs by semantic constraints (e.g., UMLS semantic types) to eliminate meaningless source/target pairs, and finally evaluates the candidate pairs. Algorithms to evaluate the source/target maps include similarity measures such as Levenshtein distance, Jaccard similarity, and TF/IDF similarity measures. Alternatively, the candidate map can be evaluated by fuzzy match algorithms such as Lucene search [55]. These algorithms associate a candidate source/candidate map with a score with which the candidate maps can be filtered using appropriate thresholds. If the target terminology has a subsumption or part-of hierarchy, the candidate maps can be further refined by choosing the most specific matches that have the best score.

6.2.3 Rule-based

The rule-based approach refers to the development and use of a computerized algorithm that would capture the high-level reasoning a human would use for the generation of maps between two terminologies. The simplest example of such a
heuristic is to consider two codes from two terminologies to be equal if these two codes have the same terms. A sophisticated rule-based algorithm could use more input parameters to define the rules for generating a map. These parameters may include: **codes** if the mapping is done between an extended national classification system (e.g. ICD-10-AM) and an international classification system (e.g. ICD-10) such that the codes in one system are related to those of another, **code terms**, the **hierarchical structure** of the classification system (chapter, block, axes, etc.), and even **similarity scores** (e.g. relevance score in ElasticSearch) that may come from a separate algorithm that uses NLP and similarity scoring.

The logic behind the rule-based approach can be implemented using various technologies. There are two aspects related to this implementation: (1) the representation of the terminologies in a machine-readable format and (2) the ruleset needs to be implemented using suitable querying techniques. Commonly used methods of representation of machine-readable formats are databases, textual tables, XML, RDF (Resource Description Framework) format [56]. Querying techniques include SQL in the case of databases, SPARQL in the case of RDF [57], a suitable programming language in the case of text files, or a combination of these techniques.

It is important to use human interventions for verification and validation of the computer-generated mapping lists (see the hybrid mapping technique explained below). The feedback received in this verification process can be used for the further refinement of the rule-based algorithm.

### 6.2.4 Machine learning

The machine-learning approach refers to the use of specific machine-learning algorithms to develop a classifier that, based on prior training on sample maps, can create new maps between two classification systems. This learning process involves several steps: data gathering, data pre-processing, the creation of a model, training of the model, evaluation of the model using a set of metrics (e.g. precision, recall, F-score and accuracy) and fine-tuning of the model for optimization [58].

The data-gathering process involves sourcing the code lists of the two terminologies that need mapping. These lists, at a minimum, should have the codes and terms associated with them. However, having more details such as the hierarchical structure (e.g. parent code, chapter, axes, etc.) would be useful in the mapping process. After sourcing the data, it may be necessary to pre-process the data to generate appropriate
inputs for the machine learning algorithms. Generally, this step involves dealing with any missing data values. Furthermore, given that code terms are written in natural language, it is necessary to use NLP techniques to identify the context of the terms (e.g. the body system to which an intervention is applicable) or to separate nouns and verbs (e.g. identification of diseases and interventions).

Creating maps between two terminologies fall into the category of “multi-label classification (binning) problem” [59] in the realm of machine learning. In this problem, multiple labels (e.g., multiple target codes) can be assigned to a problem instance (e.g., source code to be mapped). There is a wide range of algorithms suitable for multi-label classification, and it may be necessary to use a combination of these algorithms for the mapping exercise depending upon the sophistication of the input data (e.g. code and term vs code, term and hierarchy). The next important step is providing some sample training data for the machine learning algorithm. This essentially means that sample maps between the two classification systems are pre-created. If a source terminology is organized as chapters per the body system, it is appropriate to run the machine learning task on a chapter-by-chapter basis using training data for each chapter. Generally, the accuracy of the mapping becomes higher if more training data is provided. When training data is lacking, it may be necessary to run the machine-learning algorithm iteratively with feedback provided by human terminology specialists to obtain better models.

6.2.5 Evaluation of algorithm-generated maps

In computational approaches, the maps are best evaluated against previously generated human-created maps that have gone through the QA and usage validation process as described in Section 5, generally referred to as the validated ‘reference’. In computational mapping, the evaluation of generated maps against gold standards is organized into four categories in a basic evaluation matrix, as shown in Figure 2.

<table>
<thead>
<tr>
<th>Computer Generated Map Result</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>True positive (TP)</td>
<td>False negative (FN)</td>
</tr>
</tbody>
</table>
Validated reference
i.e. statement found in original source data (e.g. diagnosis as written in the medical record), manually coded, tested and validated

<table>
<thead>
<tr>
<th></th>
<th>No. correct positive maps</th>
<th>No. incorrect negative maps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>False positive (FP)</td>
<td>True negative (TN)</td>
</tr>
<tr>
<td></td>
<td>No. incorrect positive maps</td>
<td>No. correct negative maps</td>
</tr>
</tbody>
</table>

Figure 2: Basic Evaluation Matrix

While a few other metrics (precision, recall, specificity, etc.) can be used in the evaluation of the output, at a minimum, accuracy and error rate should be calculated, using the formulas shown in Figures 3 and 4, to evaluate computer-generated maps.

\[
\text{ACC} = \frac{TP + TN}{TP + TN + FN + FP} = \frac{TP + TN}{P + N}
\]

Figure 3: Formula for calculation of accuracy rate of mapping

\[
\text{ERR} = \frac{FP + FN}{TP + TN + FN + FP} = \frac{FP + FN}{P + N}
\]

Figure 4: Formula for calculation of error rate of mapping

The objective of the computational mapping exercise is to reduce the false positive (FP), and false negative (FN) maps so that the accuracy is higher and the error rate is lower.

If the human-created maps do not exist, it is important to use the services and expertise of terminology specialists for the purpose of evaluating the output generated by the computational algorithms. The results of this validation process would need to be fed back into the computational algorithms as additional training data for optimization of the algorithms. Over time, this iterative process would result in computational algorithms creating maps that are close to the gold standard quality.
6.3 Hybrid Methods

Hybrid approach refers to the use of a combination of two or more approaches defined in the previous sections. An example is the use of a rule-based approach followed by manual verification of the outputs using human terminology specialists. Also, it is possible to use more than one computational, algorithmic technique in combination for the purpose of generating the maps. For example, a rule-based mapping may be used to create the initial maps that serve as the training data for a machine-learning algorithm.

7. Case Studies

To illustrate the map development processes and techniques described in the previous sections, we present summaries of a collection of case studies. Details of most of the case studies are available as appendices of this paper.

7.1 Mappings between ICD-10 and ICD-11

Bidirectional maps between ICD-10 and ICD-11 are generated using logical statements between concepts of both classifications. These statements are partly added to the system automatically (based on the derivation of many ICD-11 concepts from ICD-10) and partly curated by the experts.

On the ICD-11 side, concepts in the ICD-11 Foundation Component have been used as the maps’ source and target.

The statements in the mapping system could be in one of the following forms:

- Equivalent concepts: This is used when the concept in ICD-10 and concept in ICD-11 are equivalent.
- Subclass / Superclass: This is used when a concept of one side is a subclass or superclass of a concept on the other side.
- Using multiple concepts in the statements (with union semantics): It is possible to use a union of multiple items when creating logical statements. For example, *Typhoid and paratyphoid fevers* in ICD-10 is equivalent to the union of *Typhoid fevers* and *Paratyphoid fevers* in ICD-11.
7.1.1 Maps

We generate two types of maps from the statements we have in the system.

- **Lead-to-one-entity map**

  This type of map aims to find one entity in the target classification. In some cases, the mapped entity is not as specific as the original, but the mapped entity semantically still includes the source entity. When there are multiple mapped concepts, the one target concept is determined by finding their common ancestor concept. This type of map is generally used in automated coding systems when it is necessary to have a single target concept.

- **Lead-to-multiple-entities map**

  This map type can list several entities from the target classification instead of trying to come up with a single entity.

7.1.2 Handling the linearizations

On the ICD-11 side, even though the actual mapping statements use entities of the Foundation Component, the generated maps use ICD-11 MMS Tabulation categories. We use the internal relations within the ICD-11 to derive the ICD-11 MMS categories from the mapping statements. (See Appendix 1 for more details.)

7.1.3 Handling ICD-11 post-coordination

In cases where it is needed, instead of defining maps to post-coordination combinations, the mapping is done to pre-coordinated foundation concepts and ICD-11’s internal post-pre-equivalence system is used to come up with a combined code.

7.2 Mappings of Rare Diseases to ICD-10 and ICD-11

OrphaNet is a European portal for rare diseases and orphan drugs. It provides a detailed classification (Orpha-codes) for coding rare diseases.
7.2.1 Mapping of Orpha-codes to ICD-10

The codes in Orpha-codes are much more granular than ICD-10 entities. Due to the difference in granularity, relatively few equivalence maps can be generated. There are approximately 500 ICD-10 codes specific to rare diseases. All other rare diseases (5000 or more) are lumped in groups where non-rare diseases can be found. To code Orpha-codes for statistical purposes while deriving ICD-10 for other health care applications, one approach is to construct a map that can be used at the time of coding. This map can also be used to aggregate data already coded with Orpha-codes in terms of ICD-10. Mapping in the reverse direction (ICD-10 to Orpha-code) is not recommended since only approximately 10% of Orpha-codes have equivalent codes in ICD-10.

To develop a map that can be used at the time of coding, linkage at the terminological level is used. In Germany's implementation, this approach was chosen to map Orpha-codes to ICD-10-GM-codes, although the methodology applies to any ICD-10 implementation. In an index file with approximately 80,000 index entries, all disease names from the Orpha-nomenclature assigned to an Orpha-code are also assigned the correct ICD-10-GM codes. This way, both codes can be assigned to a data set at the time of coding, and a map from Orpha-codes to ICD-10 is generated. A limitation of this mapping exercise was that no multilingual electronic index of ICD-10 was available. Thus, this exercise will have to be undertaken for each language if the same approach should be followed.

7.2.2 Mapping of Orpha-codes to ICD-11

For ICD-11, the situation is different. With the Foundation Component of ICD-11, an electronic index is available, and each concept within the Foundation Component is identifiable through unique identifiers. As the developers of the Orpha-nomenclature have been involved in the creation of ICD-11, almost all rare diseases should be identifiable as entries in the Foundation Component. A mapping between ICD-11 Foundation Component and the Orpha-codes should therefore comprise ideally just 1:1 equivalence maps.

7.3 Mapping of ICD-10-AM to and from ICD-10

In the biennial production of the ICD-10-AM (International Classification of Diseases, Version 10, Australian Modification), maps are created to and from its parent
classification system, ICD-10. In this process, two CSV files are the mapping outputs, where the code and term of the source classification system are presented next to the mapped code and term of the target classification.

When mapping a national classification system to the international version, some codes (four or five-digit level) in the national system do not have an equivalent code at the same level in the international classification. In such cases, the mapping is done to a code a level above in the target classification system, referred to as rolled-up mapping. From the cardinality perspective, when mapping ICD-10-AM to ICD-10, 98.75% of codes have 1:1 (equivalence or rolled-up) maps, and 0.8% of codes have 1:many maps, where one ICD-10-AM code is being mapped to multiple more specific ICD-10 codes. In the maps from ICD-10 to ICD-10-AM, 93.56% of codes have 1:1 maps, and 6.44% of codes have 1:many maps. There are no situations where m:n mappings or non-existence of maps, when mapping ICD-10 to ICD-10-AM, for two reasons: (a) ICD-10-AM is an extension of ICD-10 and (b) owing to rolled-up mapping approach. However, when mapping ICD-10-AM to ICD-10, 0.45% of the codes do not have appropriate maps. The national classification system has new ICD-10-AM codes at three-digit level (e.g. O09, Duration of Pregnancy) with no ICD-10 equivalent.

Before 2018, maps were created manually and validated by classification specialists. In 2018, maps were generated computationally with a rule-based technique, using human maps as the gold standard for verifying the computerized mapping approach. In this process, the codes, descriptors of the codes, the hierarchical structure of the classification systems and the relevance score (in Elasticsearch) were used for creating the rules for the computerized algorithm. The inputs given to the rule-based computerized algorithm were code lists and descriptors of both the source and the target classification systems. Computer-generated maps were verified against the manually created maps. In the first few iterations, the verification process helped to fine-tune the rules related to the maps. This exercise identified specific errors (less than 0.01%) that existed in the original human maps. The biennial review process of the ICD-10-AM is the motivation for continuous maintenance of these mapping files between ICD-10 and ICD-10-AM.
7.4 Mappings of SNOMED CT to ICD-10

There are two case studies for SNOMED CT to ICD-10 maps. Both are included as appendices in this paper, giving an overview of a map provided for international use and giving an overview of a map provided for national use.

The SNOMED CT-to-ICD-10 International map is a rule-based complex map produced by SNOMED International. The map is designed as a base map that Member countries may further develop to reflect national specifications and requirements. There are two types of a map in the international mapping tables with a cardinality, as defined here, of 1:1.

The SNOMED CT-to-ICD-10 (UK Edition) map produced by NHS Digital is an extended complex map that builds on the base maps provided by SNOMED International to reflect national standards, specifications and requirements. There are four types of a map in the national mapping tables and the cardinality of these, as defined here, is 1:1 and 1:many.

As discussed in Section 4, cardinality can be ambiguous unless described for a particular map. In these case studies, cardinality is defined as follows:

1:1 is either one source code to one target code or one source code to more than one target codes, all of which are required to represent the source code’s meaning.

1:many is one source code to a choice of target codes (defaults and alternatives)

The rationale for describing maps with more than one target code as 1:1 is that if one of the target codes was removed, the remaining targets would no longer represent the source’s meaning.

For examples, see the discussion on types of a map in the SNOMED CT to ICD-10 UK Edition Case Study appendix.

7.4.1 Mapping of SNOMED CT to ICD-10: international edition

The SNOMED International map is a one-directional map from SNOMED CT (source) to ICD-10 (target) produced to provide support within the specified mapping use case for SNOMED International Members and Affiliates. The map is intended to promote widespread and comparable epidemiological and statistical data without bias and is based upon the ICD-10 clinical coding rules and conventions. The SNOMED CT map to
ICD-10 (2016 version) will support organizations collecting data for statistical reporting, epidemiology, cancer, injury and other quality reporting and research. The map is deemed international and does not consider any local specifications or National Release Centre use cases (see Appendix 3).

7.4.2 Mapping of SNOMED CT to ICD-10: UK edition

This map aims to produce a reliable, consistent and reproducible link, which supports the reuse of data collected during the direct care of the patient.

The semi-automated map in one direction, from SNOMED CT to ICD-10, supports coding in England for morbidity purposes. Specifically, the ICD-10 classification is a vital component of the Admitted Patient Care dataset, Central Returns, other data sets and underpins the National Tariff payment system. It is crucial that organizations using an Electronic Patient Record (EPR) can efficiently derive ICD-10 codes for statistical and epidemiological purposes from the clinical information recorded using SNOMED CT. The derived maps support the "record once, use many times" principle and reduce manual coding. This allows classification experts to concentrate on the coding of more complex cases (see Appendix 4).

7.5 Mappings from National Classifications to ICHI

Countries with national intervention classifications have explored and conducted preliminary mapping projects to explore the feasibility of adopting ICHI and understanding the differences and similarities between the national classifications and ICHI. These projects were done to examine gaps, opportunities, and maps for international comparability. Note that in the use cases presented, ICHI is always the target classification. The mappers are required to understand the structure, content and coding guidelines used in ICHI.

The definition of cardinalities in a map depends on the national intervention classification and whether coded concepts are pre-coordinated or not. Description of map cardinality as 1:1 or 1:many may be ambiguous if the national intervention classification includes details such as devices and products used and laterality or quantification, which are separate extension codes in ICHI. In most of the mappings to
ICHI, a post-coordinated national intervention classification code is considered a single source code in a map. An expression consisting of an ICHI stem code with or without extension codes is considered a single target in a map. Thus the cardinality for mapping a national interventions classification to ICHI can be described as 1:1, 1:many or many:1, where the cardinality 1:1 means that one national intervention code (possibly with multiple attributes) may be equivalent, broader, or narrower to one ICHI stem code with or without extension code(s); the cardinality 1:many means that one national intervention code requires the conjunction of two or more ICHI stem codes with or without extension code(s) to be suitable matches (which may involve one side being narrower or broader than the other side), such as the CCI example 1.OD.89.DT-AM
Excision total, gallbladder with extraction of calculi from bile ducts using basket device endoscopic [laparoscopic] approach. In ICHI, two stem codes and an extension code would be required for equivalent meaning:
KCF.JK.AB Laparoscopic cholecystectomy/KCM.JE.AB Laparoscopic extraction of calculus from bile duct & XT06.04
Gastroenterologic basket device

In some cases, it helps enter the number of ICHI stem codes or national intervention codes instead of only using the term 'many', such as in conversion tables. We recognize that in existing national intervention classifications, the semantics of complex interventions may be a mixture of logical subsumption and composition, such as the ICD-9 CM Volume 3 example 79.31 Open reduction of fracture with internal fixation – humerus example discussed in Section 3.1. Thus the map from such a complex intervention to multiple ICHI codes may represent a has-part or an is-a relationship from the source to the target. The maps described in these case studies represent a pragmatic approach that fulfils the goal of developing a preliminary understanding of the differences and similarities between the national classifications and ICHI.

ICHI Beta-2 2018 contains approximately 7,000 interventions. Each intervention in ICHI is described in terms of three axes: Target (entity on which the Action is carried out);

4 The mapping of the Canadian Classification of Health Interventions to ICHI (described in Appendix 7) is an exception to this rule. It considers a source intervention with one or more attributes and a target expression consisting of a stem code and one or more extension codes as having the cardinality of “many”.

41
Action (deed done by an actor to the Target), and Means (processes and methods by which the Action is carried out). Additional information about an intervention can be added using extension codes.

Australia conducted two mapping exercises: ICD-9-CM to ICHI and Australian Classification of Health Interventions (ACHI) to ICHI. Canada conducted a mapping exercise mapping the Canadian Classification of Health Interventions (CCI) to ICHI. Sweden conducted one mapping exercise, mapping the National Classification of Health Care Procedures (KVÅ) to ICHI. Summaries of each mapping exercise are provided below. Further detail is found in Appendices 5 - 8.

7.5.1 Mapping from Australian Classification of Health Interventions (ACHI) to ICHI

A pilot was conducted to map ACHI to ICHI, using an algorithmic approach. This utilized a combination of computational algorithms, machine learning, human verification and manual mapping. In the initial stage, a computational algorithm using NLP was employed, followed by human verification. The resulting dataset was given as the training data to machine-learning algorithms. Then, machine-learning algorithms were used for the next iteration of mapping, followed by manual intervention for verification and mapping of up-mapped codes. It is anticipated that 80% of ACHI will be mapped to ICHI using this mechanism. The remaining 20% will be completed through human mapping, which will be used as training data for the machine-learning algorithm to increase maps' accuracy in the next iteration (see Appendix 5).

7.5.2 Mapping from ICD-9-CM to ICHI

A study was conducted to map ICD-9-CM interventions to ICHI, laying the foundation for comparing historical data to data collected via ICHI in the future. Manual mapping and verification were done by classification specialists who applied mapping rules. Links were made through 1:1, 1:many, many:1, and many:many maps, with and without the use of extension codes. Having many target intervention codes means that they collectively represent the full clinical description of the source intervention. (See Appendix 6)
7.5.3 Mapping from Canadian Classification of Health Interventions (CCI) to ICHI

The Canadian Institute for Health Information (CIHI) conducted a mapping from CCI to ICHI Beta 2018. This was conducted as part of an informal ICHI field trial. This exercise's focus included interventions most frequently reported to CIHI databases and interventions most frequently referenced in health indicators reported on by CIHI. Using the list of the top 100 interventions reported in hospital inpatient and ambulatory care and the interventions used in Canada's national health indicators, classifications specialists mapped the CCI code concepts to ICHI. Mapping rules and advice were applied to the maps as required. A preliminary analysis produced the following cardinality: \(1:1\) (70%); \(1:many\) (18%); \(many:1\) (7%); \(many:many\) (5%), where having many target intervention codes means that they collectively represent the full clinical description of the source intervention (see Appendix 7).

7.5.4 Mapping from Swedish Classification of Health Care Interventions (KVÅ) to ICHI

The National Board of Health and Welfare performed a pilot study regarding mapping two subsets of the Swedish classification of health care interventions (KVÅ) to ICHI. This pilot study was a part of the informal ICHI testing in 2018.

About 50 orthopaedic surgical interventions and 50 functioning interventions were mapped manually to ICHI by two mappers. Each KVÅ intervention was mapped to each axis in ICHI, ICHI stem code(s) and extension code(s). The degree of equivalence was explored. The cardinality for the orthopaedic surgical interventions (as defined at the beginning of Section 7.5) varied between \(1:1\) to \(1:3\), and for functioning interventions varied between \(1:1\) to \(1:6\).

There are differences in the mapping results for surgical and functioning interventions. The surgical interventions in KVÅ mainly were mapped to one ICHI stem code together with one or more extension codes. The functioning interventions in KVÅ are broader and mapped to several ICHI stem codes, and only a few extension codes were used (see Appendix 8).
7.6 Mapping from the International Classification for Nursing Practice (ICNP) to ICHI

Researchers from the UK and Canada developed mappings from the 2017 release of the ICNP to ICHI Beta 2018. This was conducted as part of the 2018 programme of informal ICHI field testing. An agreed subset of terms (the derived ICNP Community Nursing Catalogue) formed the source terminology. Two coders independently sought mappings (limited to a cardinality of one-to-one) from source ICNP terms to fully expressed target ICHI concepts (extension codes were not included). They held meetings to reach a consensus in cases of disagreement. Two senior health informatics researchers oversaw the process. The result was a set of one-to-one mappings (equivalence, broader than, and narrower than) covering 81% of source terms.

The great majority of maps (94%) were an exact match or broader (ICHI beta 3 broader than ICNP), suggesting the potential role of ICHI as an aggregating administrative classification (see Appendix 9).

7.7 Mapping of SNOMED CT to ICF

Kieft et al. [18] set out to map a Dutch nursing subset of SNOMED CT patient problems to the ICF (and Omaha System and NANDA international). For this purpose, descriptive research was performed using a unidirectional strategy based on the manual semantic mapping. This mapping method was based on ISO 18104:2014. This model's level of detail made it possible to objectify the similarities and differences between the terms to be mapped.

Out of the 119 SNOMED CT patient problems, 47 had exact 1:1 match; 9 (8%) could be partially mapped - source and target term were not the same but corresponded partially and were related. An example is the SNOMED CT term 'difficulty coping', which was linked to 'problems with cognitive flexibility & handling stress and other psychological demands' in ICF. Only one concept (1%) was mapped with the cardinality of 1:many - the source term was less detailed than the target term; 50 (42%) were mapped with the cardinality of many:1 - the source term was more detailed than the target term; and 12 (10%) were mapped 1:none - no target term was found for the source term. Since no clusters (i.e., post-coordination) are made in ICF, the many in 1:many or many:1 refer to more than one ICF codes.
Significant limitations specific to this mapping exercise (or mapping to ICF in general?) include the inability to use qualifiers and that sometimes seemingly one-to-one mappings - for example 'pain (clinical finding)’ —> 'pain (body function)’ - might come from different semantic domains, as described in paragraph 'Definitions of Terms' (Section 3).

8. Opportunities and Limitations

Often the goals of terminology mapping involve data transformations that are necessary for various health-care-related purposes. Nevertheless, the development of terminology maps may be an opportunity for improving either or both source and target terminologies. The increasing availability of logically defined terminologies also presents opportunities to create logic-based maps instead of manually created or generated using some of the automated methods discussed in Section 6.

We also need to recognize that terminology mapping is an imperfect solution to the tower-of-Babel problem in health care. Information loss, while sometimes necessary and desirable, is almost always the case. The use of terminology mapping to achieve interoperability of data is also fraught with problems. In this section, we discuss some of the opportunities and limitations of terminology mapping.

8.1 Identification of Errors and Gaps in Source and Target Terminologies

By developing a map between two terminologies, it will become apparent if one or both terminologies have gaps or limitations. This might be due to the design of the terminologies and entirely intentional. However, this finding can also be an opportunity to start discussions about, or revisions of, one or both terminologies. If, for example, the source terminology has more details in a specific area and therefore cannot be mapped 1:1 to the target terminology, and the use case for the mapped data requires the detail of the more granular terminology, this can initiate a discussion to modify the granularity of the target terminology. The evaluation of ICHI explicitly uses mappings from national intervention classifications (e.g., the Canadian Classifications of Health Interventions, see Appendix 7) to uncover potential gaps. In such cases, it would be useful for the mapping project's tooling to assist in providing feedback for augmenting the source or target terminology.
More sophisticated analysis based on the topological patterns in a map can identify a potential error or missing intermediate concepts in either source or target terminologies. For example, suppose a mapping establishes that concepts A1 and B1 from terminology 1 are equivalent to concepts A2 and B2 from terminology 2, respectively, and concept X in terminology 1 is an intermediate concept broader than B1 but narrower than A1. X has no equivalent concept in terminology 2. In that case, X's position between A1 and B1 is either an error or X is a potential intermediate concept between A2 and B2 that can be added to terminology 2. He et al. [38, 60] applied this type of analysis to mappings from eight terminologies in UMLS to National Cancer Institute Thesaurus to identify hundreds of potential concepts for addition to the source terminologies.

8.2 The Use of Ontology and Automated Reasoning

While speculative as a practical mapping method today, the emergence of ontologies that encode an increasing amount of health-care knowledge using Description Logics (DL) [25, 61] portends the opportunity to use formally defined knowledge in reasoning about concepts. This is already demonstrated in ontology authoring and maintenance, where determining all subsumption relationships across very large terminologies can be guaranteed using computational reasoners. Correspondingly, new concepts can be correctly positioned in an ontology through logical reasoning, while many “inferred” concepts can be algorithmically generated using similar reasoning to enrich an ontology. The principle, simplistically, is to extend computational reasoning across ontologies to support mapping between them. This has been demonstrated experimentally by Andronache et al. [62], who ontologized ICF and concepts in assessment instruments and used DL reasoning to link instruments to ICF.

As part of the experimentation in the development of ICD-11, a collaboration between WHO and IHTSDO applied this ontological approach to the alignment of the ICD-11 Foundation Component and SNOMED CT. To semantically anchor the Foundation Component, creating a Common Ontology as a subset of SNOMED was proposed. Since SNOMED CT is based on description logic, it can support reasoning. However, many ICD concepts, and thus many of those in the Foundation Component, such as Hypertension excluding Pre-eclampsia, have exclusions that involve negation (all hypertension cases but NOT those with pre-eclampsia) which is either not supported or inefficient in most description logics. To accommodate this mismatch between SNOMED CT and ICD, we
invoke query logic used in SQL-style operations to link Foundation Component concepts to the Common Ontology [13, 63, 64]. This approach was demonstrated for the cardiovascular chapter of the ICD. The Common Ontology remains to be completed.

The application of this reasoning infrastructure to mapping leverages the logical relationships defined among concepts. Given that the proposed Common Ontology is a subset of both SNOMED CT and ICD Foundation Component, any concept in either terminology fully definable in terms of the Common Ontology will automatically be placed in hierarchies of the other. Other concepts cannot be fully defined, either because they are primitive concepts or because they are ICD concepts with exclusions and need to be handled using other methods. Suppose the mapping can be formulated in logical terms like the mapping statements relating ICD-10 and ICD-11 concepts. In that case, the same kind of algorithms used in the ICD-10/ICD-11 mappings can generate mappings from SNOMED CT concepts to any ICD-11 linearizations (e.g., the MMS Tabulation).

Description logics can be used to generate maps and the application of maps if target concepts can be fully defined (i.e., with necessary and sufficient conditions) using source concepts and relationships. In this mapping approach, the codes from a specific case define the correct target code precisely. In place of rules in rule systems, definitional logic relationships enable description logic reasoning to invoke co-occurring conditions or findings to establish the target code.

8.3 Limitations on the Creation and Maintenance of Mappings

Maps are created between two terminologies, each of which has its intended usage, scope, approach of structuring terminology, and history of creation and evolution. Naturally, these differences will impose limitations on the mappings that are possible to construct. The meaning of a concept in a terminology is defined not only within the concept itself but also by its location in the terminology’s hierarchies and other characteristics. For example, the dagger-and-asterisk system in ICD-10 constrains how specific codes can be used. Another issue is the possibility of implicit context in using a particular term that a map does not fully capture, leading to incorrect mapped data. For
example, when a generic code for a TNM grade\(^5\) is used in a local pathology report template for a cancer, which, depending on the context, may be breast or prostate cancer, for example, the local code should be mapped to different cancer-specific codes. While creating and revising a map, all these aspects must be considered. Suppose the structure of and constraints on one terminology are too different from those of the other. In that case, the map might be challenging to create and maintain, even though individual pairs of source and target concepts might seem to be quite similar. It also means that a map will often be approximate and dependent on the developers' judgment and users of the map.

8.4 Information Loss and Interoperability Problems with Mapped Data

Terminology mappings can only be applied with an understanding of the limitations that are recognized and documented as part of their development and maintenance. Because of the inherent differences between two terminologies, information loss using maps on previously coded data is the norm rather than the exception. Sometimes this information loss is necessary and desirable, as in mappings from detailed codes necessary for clinical care to aggregated codes for statistical analysis. Otherwise, the secondary uses of coded data need to take this information loss into account. In general, only if the map is used at the time of coding when the coder might have the opportunity to code a more precise concept (e.g., using extension codes), and a knowledgeable user is available to verify the chosen concepts in both terminologies can such information loss be minimized.

States et al. [65] identify additional problems in using terminology maps to achieve interoperability of data. The usage scenario involves maps from two local terminologies to the same standard terminology (e.g., ICD). Characteristics of the maps directly impact whether data sets created using these maps are comparable. For example, one map may link a source concept to the single closest target concept, but the second map adds relation labels such as *exact, narrower than, broader than*. When no exact match is

\(^5\) The TNM Staging System for grading cancer is based on the size and extent of the tumour (T), the number of nearby lymph nodes that have cancer (N), and the presence of metastasis (M).
available, one map may map a source concept to a broader concept in the target terminology. In contrast, the second map may require that a user chooses among a set of narrower concepts, or it may default to a narrower concept that occurs most frequently in the data set. The two maps may also map the "same thing" to different codes in the target terminology. For example, the concept of weight may be mapped to LOINC 29463-7 Body weight or LOINC 3141-9 Body weight measured. Finally, differences in the granularity levels in the source terminologies compared to that of target terminologies may lead to the same concept in the source terminologies being mapped to different target codes. For example, the concept of Infiltrating duct carcinoma of a lower inner quadrant of the breast has two possible targets in SNOMED CT: Malignant neoplasm of breast lower inner quadrant or Infiltrating duct carcinoma of breast. Thus, the same observation recorded in the source terminologies can be mapped to different target terminology codes.

9. References


5. Friedman C, Rindflesch TC, Corn M. Natural language processing: state of the art and prospects for significant progress, a workshop sponsored by the National


50. Vikstom A, Skaner Y, Strender L-E, Nilsson GH. Mapping the categories of the Swedish primary health care version of ICD-10 to SNOMED CT concepts: rule development and intercoder reliability in a mapping trial. BMC medical


54. cTAKES™ A. Apache cTAKES™ is a natural language processing system for extraction of information from electronic medical record clinical free text.: The Apache Software Foundation. Available from: http://ctakes.apache.org/whycTAKES.html.


10. Glossary

The definitions of terms given in this glossary are specific to their usage in this paper. Some terms (e.g., cardinality, false positive, false negative) have more general meaning outside this context.

**Cardinality**: The cardinalities of a terminology map, written as n:m, is the number of source entities (n) and the number of target entities (m) in the map. The nature of entities being counted needs to be defined precisely in the mapping.

**Classification**: A type of terminology where (1) each classification entity has only one parent; (2) sibling entities should be mutually exclusive; and (3) the union of sibling entities should be semantically equivalent to the parent entity.

**Clinical terminology**: A type of terminology used directly or indirectly to describe health conditions and health care activities.

**Description Logics**: A family of logic-based knowledge representation formalisms that can be used to model domains as ontologies. An ontology consists of concepts (aka classes), roles (aka properties, relationships), individuals (aka instances), and relationship among them. With appropriate definitions (logic axioms), automated reasoners can check the consistency of an ontology, compute subsumption relationships in a taxonomy, and categorise individuals into classes of which they are instances.

**Dual independent mapping**: The process of having two (or more) mapping experts independently create maps for the same set of source and target concepts. The purpose of the process is to uncover potentially ambiguity in the relationships between source and target concepts.

**Extension code**: In a terminology that allows post-coordination of concepts, extension codes are values that can be used to qualify a stem code to form a specialization of the concept represented by the stem code.

**False-negative**: Given a gold standard of accurate maps from one terminology to another, the number of correct maps not generated by a mapping algorithm.

**False-positive**: Given a gold standard of accurate maps from one terminology to another, the number of incorrect maps generated by a mapping algorithm.

**Foundation Component**: A knowledge base that includes all knowledge content relevant to ICD in its various formulations. A formal Content Model specifies the
information that should be specified for entities in the Foundation Component. Different variants of ICD, called linearizations, can be generated algorithmically from the Foundation Component.

Gold standard: In the context of measuring the performance of information systems, a gold standard is a set of cases that represent the “correct” results for the given inputs. Systems that generate terminology maps automatically are often evaluated against “gold standard” maps that experts developed manually through a consensus process.

Linearization: In the information infrastructure used to develop ICD-11, linearizations are variant versions of ICD-11 designed for specific use cases (e.g., primary care, dermatology speciality). They are programmatically generated from the knowledge content and specifications stored in the Foundation Component. The ICD-11 Mortality and Morbidity Statistics (MMS) tabulation, for example, is a linearization designed to be the successor to ICD-10.

Machine learning: A branch of computer science that studies how to develop systems that can improve their performance on some tasks after data exposure. In the context of using machine learning to generate candidate terminology maps, a system is exposed to examples of accurate maps and is expected to find new candidate maps.

Map: A map is either an individual map (i.e., “index from one term to another”) or a map set (i.e., a “group of individual maps used to convert a range of entries from source to target code”). (Adapted from [14])

Mortality and Morbidity Statistics (MMS) tabulation: The successor to ICD-10 in ICD-11. It is one of the linearizations in ICD-11.

Multi-label classification: The problem of assigning one or more labels (e.g., multiple target codes) to a problem instance (e.g., source code to be mapped).

Natural language processing: A branch of computer science concerned with building systems that can process natural language data to support various tasks (e.g., speech recognition, natural language understanding). In terminology mapping, natural language processing techniques are applied to process the textual descriptions of concepts before matching them as candidate maps.

Ontology: A logic-based representational artefact that formalizes the entities and relationships in a domain of discourse [22, 23].
**Post-coordination**: The ability to combine multiple concepts to represent concepts that do not have fixed identifiers in the terminology. The method to form new concepts may use logical operations like conjunction, disjunction, or negation, relations such as *associated with*, or it may involve multiple axes of descriptors (aka *extension codes*) that can be chosen to modify an existing concept (aka *stem code*).

**Pre-coordination**: Fixed enumeration of all possible concepts in a terminology.

**Recoding**: The process of using primary source materials to code in a second terminology. Contrast with the use of terminology mapping to derive coded data in a second terminology from previously coded data.

**Rolled-up mapping**: In mapping a national classification to the international version, when some codes in the national system do not have an equivalent code at the same level in the international classification, the mapping is done to a code a level above in the target classification.

**Semantic domain**: A set of entities that are of the same kind.

**Semantic equivalence**: Two concepts are semantically equivalent if the sets of things referenced by them are the same.

**Semantic match**: The process of comparing two terms based on how their meanings are related to each other.

**Similarity metric**: A real-valued function that quantifies the similarity between two objects.

**Stem code**: In a terminology that allows *post-coordination*, the code that is qualified by one or more *extension codes*. The concept represented by the stem code (e.g., *infection*) is specialized along the post-coordination axes of the *extension codes* (e.g., *an infection caused by bacteria*).

**Terminology**: A set of designations (i.e., representations of concepts by signs which denotes them) belonging to one particular language. In this paper’s context, the particular language is the language for describing entities and relationships in the health care domain. (Adapted from [15])

**Terminology mapping**: The process of defining a relationship between concepts in one coding system to concepts in another coding system, in accordance with a documented rationale, for a given purpose. (Adapted from [14])
11. Appendices

Appendix 1: ICD-10/ICD-11 Mapping Tools and Its Output Tabulation

Author: Can Çelik, WHO

The mappings between ICD-10 and ICD-11 are specified using a collection of logical mapping statements that state equivalences or class-subclass relations between the entities of two classifications. Developers use a mapping tool to create the mapping statements.

The Mapping Tool

The mapping tool's core function is to collect information about the mapping relations between the concepts of the ICD-11 Foundation and ICD-10. The tool allows us to state equivalences as well as subclass and superclass relations between the classifications.

Relation Types

The following examples illustrate various types of information that we could enter into the tool:

Equivalence (one-to-one mapping)

This is the simplest and the most common type of relationship in which the concepts on both ends are equivalent. The tool uses the $\equiv$ sign to show equivalences.

Example:

<table>
<thead>
<tr>
<th>Chronic myelomonocytic leukaemia</th>
<th>$\equiv$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C93.1 Chronic myelomonocytic leukaemia</td>
<td></td>
</tr>
</tbody>
</table>

Edit
Subclass and Superclass Relations

In some instances, there is no exact equivalence in the other classification, but we know where our concept could be included. In such cases, we use subclass or superclass relations depending on which concept is broader.

The tool uses the ⊃ sign to show superclasses (i.e. the concept above is broader) and ⊂ for the subclasses (concept above is narrower). The software will provide a tooltip when the mouse hovers on the icon.

In this example, "Other specified disorders of pigmentation" is a subclass of Disorders of skin colour.

Relations Using a Group of Entities

In some instances, our relation may involve multiple entities from the classifications. These types of relations are most likely to use subclass or superclass relation.

In this example, "Anaemias" in ICD-11 is a broader concept than the union of all three ICD-10 concepts.

Using the Tool for Editing the Mapping Statements

Basic Editing

The mapping tool main screen consists of 3 parts:
● On the left, we have ICD-11 Foundation
● On the right, we have ICD-10.
● At the centre, are displayed mapping statements. The tool shows the mapping statements made for the entity that have been clicked, which could be an entity on any one of the classifications.

Editing starts with selecting an entity from ICD-11 or ICD-10. Search or advanced search features may be used to move to the desired entity and browsing through the hierarchy.

Clicking on the "Add New Mapping Statement" adds a new mapping statement. Once it is created, the user could click the Edit button and drag more entities from both classifications to the statement. The user could set or change the type of relationship as well (i.e. Equivalent, Subclass, Superclass)

Crosswalk

There is no need to have a mapping statement for every entity in the classification. The system finds a derived mapping for the cases where we do not have any such mapping statements defined.
The system shows this under the **Crosswalk** section. Clicking on an item at the ICD-11 hierarchy will show where it would be mapped in ICD-10, and when clicking an entity at the ICD-10 hierarchy, it will show where it would be mapped in ICD-11. More information on how this is done can be found in the "Mapping Tool Tabulations" section.

Checking Consistency

The mapping statements are logic statements that state equivalences or class-subclass relations between the entities of two classifications. When we join these relations with the class-subclass relations within the classifications and the mutual exclusion rules of ICD-10 and ICD-11 linearizations, we end up with a collection of logical statements that sometimes conflict with each other.

The consistency checking mechanism performs a consistency check around the area that is being browsed. If it finds any issues, it will give entities that take part in the inconsistency.

Mapping Tool Tabulations

The tool generates two types of mapping tabulations from the statements collected by the system.

- **Lead-to-one entity map**
  - In this map, we always try to find one entity in the other classification. In some cases, the mapped entity is not as detailed as the original, but the mapped entity semantically includes the entity (i.e. is a superset). We do this by leading to the common ancestor concept when there are multiple mapped concepts. This type of map is generally used in automated coding systems.

- **Lead-to-multiple entities**
  - When there is no one-to-one relation, this map could list several entities from the other classification instead of trying to lead to the common ancestor.

Since there are both ICD-11 (Joint Linearization) to ICD-10 and ICD-10 to ICD-11 (Joint Linearization) outputs, this makes four outputs from our system:

- ICD-10 to ICD-11 Joint Linearization for Mortality and Morbidity Statistics (lead-to-one entity)
- ICD-10 to ICD-11 Joint Linearization for Mortality and Morbidity Statistics (provide multiple categories as necessary)
- ICD-11 Joint Linearization for Mortality and Morbidity Statistics to ICD-10 (lead-to-one entity)
- ICD-11 Joint Linearization for Mortality and Morbidity Statistics to ICD-10 (provide multiple categories as necessary)

The Rules Used in Generating the Tabulations for single code map

1. There is a one-to-one Equivalent relation
2. If there is a mapping statement with an equivalent relation, the map points to the equivalent entity.
3. There is a relation which states A1 is a subclass of B1
4. A will be mapped to B, but the mapping will not be used in the reverse direction

5. If there is no equivalent or subclass mapping at the existing entity, the parent's rules are applied and the mapped concept of the parent in the tabulations is used. (Note: If the source classification (Classification A) is ICD-11, we use the linearization parent in this rule.)

Also, note that superclass relations are **NOT** used (i.e. if A1 parent is a superclass if B1).
6. If the concept in classification A is mapped to multiple concepts in classification B
Such as $A_1 \equiv (B_1, B_2, ...) \text{ or } A_1 \subset (B_1, B_2, ...)$
we map the entity to the common ancestor of if $(B_1, B_2 ...)$ in classification B

---

The Rules Used in Generating the Tabulations for Multiple Code Map

1. If the entity has already a map, has a mapping statement, all values are listed in that
   mapping statement. This is done regardless of the relationship type.
2. If there is no mapping statement at the entity, the maps of the children are listed.
3. If the entity does not have any children, the parent’s map is used.
4. If the descendant maps contradict the current map, then raise the issue in the output.

Handling the Linearizations:

The ICD-11 mapping software uses the ICD-11 Foundation when the user defines the mapping statements. However, our mapping tabulations are based on the Joint Linearization for Mortality and Morbidity Statistics. Because of this, the tool needs to make necessary conversions while generating the outputs.

- When the source classification is ICD-11:
  - Only use the entities that are the linearization are used. Other than that, the tabulation generation rules stated above are used for the non-residual entities.
  - Since the residuals do not exist in the foundation, no mapping statements are coming from this system.
    - For unspecified residuals, the parent is used to find the correct mapping
    - For the “other specified” residuals: if the parent has an equivalence mapping and there exists an “other specified” (.8) child at the mapped entity, a mapping to it is generated.

- When the source is ICD-10:
  - The mapping foundation concept is identified using the rules listed above, and if the resulting entity is in the linearization, then it is used. Otherwise, it is rolled up until an entity is reached in the linearization and then provided as the result.
  - Since ICD-10 residuals are available in the editing platform, do not no special processing is done for them but use the mapping statements entered in the system.
Appendix 2: International Classification of Disease (ICD-10) to International Classification of Disease Australia Modification (ICD-10-AM)

Author: Anupama Ginige, Western Sydney University, Australia

Source Terminology
International Classification of Diseases (ICD-10)

Target Terminology
International Classification of Diseases Australia Modification (ICD-10-AM)

Purpose of this Mapping (use case supported)
- Support longitudinal morbidity statistics
- Interoperability for International epidemiological reporting

Example of Usage
- Support the comparison of data coded with ICD-10-AM with a set of data coded with ICD-10 in the future.

What the Maps Are
The maps will be a linkage from an international classification, ICD-10, to a target national classification, ICD-10-AM, in one direction. The disease description is matched to one or more ICD-10-AM code descriptions. Coding rules of inclusion terms and exclusions are used as necessary.

Inputs
Source ICD-10 10th edition codes and their full description and inclusions and optional fifth-digit level code are used.

ICD-10-AM 11th edition is used that contains complete code, descriptions and exclusions.
Outputs

Mappings from ICD-10 containing codes and description to ICD-10-AM codes and descriptions as well as comments.

Types of Map

There are two types of maps:

1. Simple map – A single ICD-10 maps to a single ICD-10-AM code with accurate semantic representation.
2. Complex map – A single ICD-10 code maps to multiple ICD-10-AM codes where each of the target ICD-10-AM codes is more specific than the source ICD-10 code.

The Cardinality of a simple map is one-to-one. A Complex map is one-to-many where one ICD-10 code maps to multiple ICD-10-AM codes.

Mapping Toolkit

- ICD-10 10th Edition
- ICD-10-AM 11th Edition
- ICD-10-AM Coding Guidelines [1]
- Indexed ICD-10-AM codes in Elastic Search
- Human mapping file of ICD-10 to ICD-10-AM

Mapping Methodology

Most of the mappings are exact description match, barring American to Australian spelling, synonyms and slight difference in special character encodings and others. Full-text search using "term frequency - inverted document frequency" (tf-idf) [2][3] is an excellent technique to find mappings automatically. This technique is built into Elastic Search [4] with further improvements.

Starting with an ICD-10 description, each word in the sentence is used to find a matching ICD-10-AM description. The greater number of words match, confidence in finding a correct match increases (term frequency). However, this technique in isolation is insufficient in reducing the scope of search. Inverted documents frequency allows matching of words not occurring commonly to ICD-10-AM. The idea is to eliminate commonly occurring words and let these 'rare' words find the correct ICD-10-AM
mappings. Tf-idf produces a vector for each ICD-10 and ICD-10-AM descriptions. During a search, a similarity index is calculated between each ICD-10 to all ICD-10-AM, and the top 10 scores are used for analysis. This is repeated for all ICD-10 codes.

From the output of the Elastic Search, further processes are taken to pick the best mapping. If the top result has the same code, that mapping is taken immediately. The chance of getting correct mapping when both code and description matched exactly are very high.

There may be instances where there are two exact ICD-10-AM description matches. Only the one with the matching code is taken. The other code is not to be mapped as it is used in conjunction with another primary diagnosis.

Afterwards, synonyms, American to Australian spelling, apostrophes, dashes, and stop words are removed, just like what had been done with indexed ICD-10-AM. This is done so that string comparison can be made manually. If the top result has matching code, it is selected as correct mapping.

Many times, Elastic Search returns children of a correct mapping because of tf-idf algorithm. A combination of its ICD-10-AM parent information is used to decide if this child is indeed a child of a correct mapping.

Lastly, a 3-grams technique in a separate ICD-10-AM index is used, and a search is performed. This takes care of slight spelling differences.

All mappings that could not be found are examined to see if the algorithm can be improved. The resulting mapping is then compared to a human mapped file for validation.

Figure 1: Summary of ICD-10 to ICD-10-AM Algorithm
Example of Mappings

Table 1 One to one mapping

<table>
<thead>
<tr>
<th>ICD-10 Code</th>
<th>Description</th>
<th>ICD-10-AM Code</th>
<th>Description</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>A00.9</td>
<td>Cholera, unspecified</td>
<td>A00.9</td>
<td>Cholera, unspecified</td>
<td>Direct code and description match</td>
</tr>
<tr>
<td>F00.9</td>
<td>Dementia in Alzheimer disease, unspecified</td>
<td>F00.9</td>
<td>Dementia in Alzheimer’s disease, unspecified (G30.9+)</td>
<td>Slight change in the ICD-10-AM description as it includes dagger code</td>
</tr>
<tr>
<td>U84.8</td>
<td>Resistance to other specified antimicrobial drug</td>
<td>Z06.78</td>
<td>Resistance to other specified antimicrobial drug</td>
<td>Different code.</td>
</tr>
<tr>
<td>T36.1</td>
<td>Poisoning: Cefalosporins and other beta-lactam antibiotics</td>
<td>T36.1</td>
<td>Cephalosporins and other beta-lactam antibiotics</td>
<td>Drug spelling difference solved with n-gram technique</td>
</tr>
</tbody>
</table>

Table 2 One to Many mapping

<table>
<thead>
<tr>
<th>ICD-10 Code</th>
<th>Description</th>
<th>ICD-10-AM Code</th>
<th>Description</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>A41.5</td>
<td>Sepsis due to other Gram-negative organisms</td>
<td>A41.50</td>
<td>Sepsis due to unspecified Gram-negative organisms</td>
<td>ICD-10-AM expanded to include unspecified</td>
</tr>
<tr>
<td>Code</td>
<td>Description</td>
<td>Special Characters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------------------------------</td>
<td>--------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A41.51</td>
<td>Sepsis due to <em>Escherichia coli</em> [E. Coli]</td>
<td>and specified bacteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A41.52</td>
<td>Sepsis due to <em>Pseudomonas</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A41.58</td>
<td>Sepsis due to other Gram-negative organisms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C88.0</td>
<td>Waldenström macroglobulinaemia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C88.00</td>
<td>Waldenstrom macroglobulinaemia, without mention of remission</td>
<td>Special character solved by normalizing special characters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C88.01</td>
<td>Waldenstrom macroglobulinaemia, in remission</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Quality Assurance/Validation**

The computer-generated output is checked against the human-mapped file by experienced clinical coders.

**Drivers for Maintenance of Existing Maps**

- ICD-10 update cycle
- ICD-10-AM revision cycle

**Competencies and Skill Set for Map Specialists**

- A highly experienced clinical classification specialist will carry out the verification and validation of the maps.
Experts in programming techniques with some background knowledge in clinical classification will develop the computerized algorithms.

Reference


Appendix 3: SNOMED International, mapping SNOMED CT to ICD-10

Author: Donna Morgan, Map Terminologist SNOMED International

Source terminology
SNOMED CT

Target terminology
ICD-10 version 2016

This mapping is a tabular, knowledge-based cross-link from SNOMED CT to ICD-10 in which the most accurate ICD-10 target code or codes that best represents the SNOMED CT concept are linked. The map is a link directed from the source SNOMED CT concept to the target ICD-10 statistical classification.

The map is deemed international, and for use by all SNOMED International member countries, so it does not consider any local specifications or National Release Centre use cases.

Purpose of mapping
- The purpose of creating a map from SNOMED CT to ICD-10 is to provide a semi-automated coding of ICD-10 statistical classification data. This is intended for use within the clinical record and encoded in SNOMED CT
- The encoded data will be utilised in clinical registries and health records and will also be used in diagnosis groupers
- Re-use of clinical data for additional statistical purposes
- Rapid submission and response to national reporting requirements
- Saving time and improving efficiency for the coding professional
- Improved accuracy and reproducibility of code mapping
- Promotion of widespread, comparable, epidemiological and statistical data

The map is intended to provide support within the specified mapping use case for SNOMED CT International Members and Affiliates that:
- have deployed, or are deploying, SNOMED CT in clinical information systems;
- use the 2016 edition of ICD-10 in systems for statistical reporting, epidemiology, cancer, injury and other registries, quality reporting, safety reporting and research.

This map is also intended to support use by WHO Collaborating Centers in interested countries where SNOMED CT is deployed or is being deployed.

What the maps are:

The SNOMED International map produced by the Mapping Service Team is a direct link from a terminology to a classification that reflects the SNOMED CT concept’s clinical meaning. The mapping considers all concept relationships, qualifiers, and attributes created when authoring the concept. The maps are a one-directional link from a SNOMED CT concept to one or more ICD-10 target codes. The goal of the mapping process is to identify the meaning of a SNOMED CT concept, determine the best location of that concept in the ICD-10 semantic space as identified by one or more ICD-10 classification codes, and to create a link between the SNOMED CT concept identifier and the correct ICD-10 code(s).

Inputs:
- Individual SNOMED CT concept (source), including their definition and attributes; ICD-10 codes (target)

Outputs:
- Individual SNOMED CT concepts with associated ICD-10 codes and mapping attributes (associated map advice)

Relationships between inputs and outputs:

All pre-coordinated concepts issued by SNOMED International within the current international release of SNOMED CT with active status within the following SNOMED CT domains may be mapped:

- Clinical finding (disorders and findings) Concept.id 404684003 and descendants
- Event Concept.id 272379006 and descendants
- Situation with explicit context Concept.id 243796009 and descendants excluding Procedure with explicit context Concept.id 129125009 and its descendants
All chapters of ICD-10 are considered within scope for this map. The Morphology of Neoplasms, which is a nomenclature of codes designed for use in conjunction with Chapter 11 Neoplasms, is not within scope for this map. We do produce and distribute a map from SNOMED CT to the International Classification of Diseases for Oncology (ICD-O).

The granularity and purpose of ICD-10 is different from that of SNOMED CT. SNOMED CT is a comprehensive reference terminology that supports both general and highly specific concepts. Each concept is defined by a set of attribute-value pairs (relationships), which uniquely define it distinct from all other concepts. SNOMED CT supports a model of meaning that specifies correct attributes and value sets for each meaning domain.

ICD-10 is a classification of diseases and related health problems with a granularity of definition chosen to provide utility for epidemiology and statistical reporting of mortality and morbidity. ICD-10 was created to classify a clinical concept by defining the classes (or 'buckets' of meaning) which contain the concept within the universe of ICD-10 classes.

Only domains of SNOMED CT, which overlap in meaning with those of ICD-10, will be mapped. Due to differences in granularity, purpose and rubrics, assignment of a mapping equivalence between the SNOMED CT source and ICD-10 target code is usually not appropriate. Instead, the MAP will link a SNOMED CT source concept to the ICD-10 code, which contains the SNOMED CT concept's meaning as conceptualized by ICD-10.

Types of map

The cardinality of the map is $1:1$, and there are two types of map to represent the concepts and its target best:

Map type 1 (simple map)

Links a single SNOMED CT concept to a single classification code to represent the clinical meaning of the concept.

Map Type 2 (complex map)

Links a single SNOMED CT concept to a combination of classification codes which collectively represents the meaning of the SNOMED CT concept.
Mapping attributes:

**mapTarget:** The *target code* in the mapping terminology browser to link the concept to.

**mapGroup:** an integer assigned to each set of Map members, which are coordinated to specify one target ICD-10 code for the map, or the null map if the source concept does not require an additional ICD-10 code for proper classification. Each mapGroup collates and orders the rules, which are sequentially evaluated to yield at most a single target code. The first mapGroup designates the set of records used to specify the first (primary) target code. The second mapGroup identifies the set of data records for the second target code. These are repeated as required to specify a complete set of mapping target codes.

**mapRule:** A machine-processable truth statement created to evaluate to “true” or “false” at run-time, which determines whether the Map member should be validated as the correct link to the associated map target for the mapGroup being evaluated.

**mapAdvice:** human-readable textual advice that a software vendor may employ to inform the clinician user or the classification expert during a semi-automated mapping session. a summary statement of the mapRule logic in readable terms for the clinical user

- supplementary metadata guidance intended to clarify the map for the coding professional. Metadata advice supported in the MAP includes:
  - FIFTH CHARACTER REQUIRED TO SPECIFY THE SITE FURTHER

The usual scenario is where a SNOMED CT concept that is general and does not specify the site maps to a target code in ICD-10 Chapter XIII (M) that is not site-specific and requires a 5th character. Therefore in this scenario, it is correct to add the 5th character .9 Site unspecified.

In certain circumstances, for a site-specific SNOMED CT concept where the SNOMED CT site cannot be matched explicitly to a 5th character site in ICD-10, the Mapping Specialist can complete the mapping to the 4th character only. To complete the code with a 5th character requires additional clinical information. In these cases, it is more useful to allow the user who has access to the clinical information to assign the 5th character rather than the Mapping Specialist assigning either the .8 Other or .9 Site unspecified 5th character.

Example
279039007 Low back pain (finding)

Map to M54.5 Low back pain plus map advice FIFTH CHARACTER IS REQUIRED TO SPECIFY THE SITE FURTHER. This is because the fifth characters provided to specify the part of the spine involved in ICD-10 Chapter XIII, M40-M54 (except M50 and M51) Dorsopathies, do not include ‘low back’. The mapper would have to choose from one of the fifth characters provided, and that choice constitutes a clinical judgment (sites are M54.56 Lumbar region, M54.57 Lumbosacral region, M54.58 Sacral and sacrococcygeal region, and M54.59 Site unspecified).

- MAPPED FOLLOWING SNOMED GUIDANCE

When mapping a SNOMED CT concept, if SNOMED CT represents a different meaning than ICD-10, the Mapping Specialist will assign a target code based on the information given within SNOMED CT along with map advice MAPPED FOLLOWING SNOMED GUIDANCE.

- MAPPED FOLLOWING WHO GUIDANCE

Mapping is to be performed according to the WHO guidelines, even if those guidelines differ from country-specific guidelines. WHO guidelines have the ultimate authority in the outcome of codes.

Example

In the United Kingdom, tobacco use of any amount is coded to F17.1, Mental and behavioral disorders due to use of tobacco, harmful use. However, the WHO index clearly leads to code Z72.0.

For these types of scenarios, WHO indexing is followed, not country-specific guidelines or norms. The map advice MAPPED FOLLOWING WHO GUIDANCE is not required.

- POSSIBLE REQUIREMENT FOR ADDITIONAL CODE TO FULLY DESCRIBE DISEASE OR CONDITION

The map advice of POSSIBLE REQUIREMENT FOR ADDITIONAL CODE TO FULLY DESCRIBE DISEASE OR CONDITION is to be added when the SNOMED concept maps to an ICD-10 code that has the Tabular instruction to “Use additional code, if desired, to identify…”

Example

19399000 |Acute exudative otitis media (disorder)|
ICD-10 Index

Otitis
- media
  - - acute or subacute
  - - - exudative H65.1

ICD-10 Tabular List

H65 Nonsuppurative otitis media
  Includes: with myringitis

Use additional membrane code (H72), if desired, to identify the presence of perforated tympanic membrane.

Map: H65.1 Other acute non-suppurative otitis media with the map advice POSSIBLE REQUIREMENT FOR ADDITIONAL CODE TO FULLY DESCRIBE THE DISEASE OR CONDITION.

- POSSIBLE REQUIREMENT FOR AN EXTERNAL CAUSE CODE

An external cause code from Chapter XX is used in combination with a code from another chapter to add information, especially if it instructs to add such an additional code. The external cause code should always be sequenced AFTER the disease chapter code.

An external cause code from Chapter XX External causes of morbidity and mortality V01-Y98 is designed for the classification of:

- external events and circumstances which are the cause of injury (includes transport accidents)
- poisoning
- other adverse effects

If a Mapping Specialist assigns an ICD-10 code which requires an external cause code to provide the reason why it happened, and the information is not available in the description for the concept to be mapped (for example, to classify the circumstances of an injury) then the Mapping Specialist must add the following map advice: POSSIBLE REQUIREMENT FOR AN EXTERNAL CAUSE CODE
**EXCEPTIONS:** There are instances where an instruction is provided at a Block or Chapter Level in ICD-10, but good judgment would indicate that the advice is not relevant to every code in that block or chapter.

Examples:

- It is not necessary to assign the external cause map advice, even if directed to do so at ICD-10 block level if it is clearly not relevant, for example:
  - Do not use external cause advice to identify a drug for every Diabetes Mellitus (E10-E14) not specified as drug-induced.
  - Do not add external cause advice if the concept is described as idiopathic.
  - If the cause is coded, do not use external cause advice. For example, do not use this advice if the concept is a kidney disorder **due to diabetes** that includes a target code in the block Renal failure (N17-N19) that has the advice at block level *Use additional external cause code (Chapter XX), if desired, to identify external agent at the Block level.*

- **POSSIBLE REQUIREMENT FOR CAUSATIVE AGENT CODE**
  
  Use this map advice when the SNOMED concept maps to an ICD-10 code that has the following Tabular instructions:
  
  - Use additional code, if desired, to identify infectious agent or disease
  - Use additional code (B95-B97), if desired, to identify infectious agent
  - Use additional code (B95-B96), if desired, to identify bacterial agent

- **POSSIBLE REQUIREMENT FOR MORPHOLOGY CODE**

  According to this principle, all source concepts representing neoplastic disorders in code ranges (C00-D48) will be mapped. Morphology mapping with ICD-O is out of scope for the MAP. The Mapping Specialist will record an advice note to denote that a morphology code may be required for completeness.

- **POSSIBLE REQUIREMENT FOR PLACE OF OCCURRENCE**

  For codes **W00-Y34** ICD-10 provides characters to identify the place of occurrence of the external causes, where relevant.

  The fourth character Place of Occurrence codes are not added to the Map, but instead, the Mapping Specialist selects Map advice **POSSIBLE REQUIREMENT TO IDENTIFY PLACE OF OCCURRENCE.**
Exceptions: Some codes within (W00-Y34) have their own fourth character, not related to the place of occurrence. Transport accidents (V01-V99) have their own fourth character classification. Map to the fourth character for the following:

- Transport accidents (V01-V99)
- Victim of Earthquake (X34)
- Exposure to unspecified factor (X59)
- Neglect and abandonment (Y06)
- Other maltreatment (Y07)
- Legal intervention (Y35)
- Operations of war (Y36)

Example

218164000 |Accident caused by electric current (event)|

W87 Exposure to unspecified electric current and add two pieces of map advice:

1. Possible requirement to identify place of occurrence
2. This is an external cause code for use in the secondary position
   - THIS CODE IS NOT TO BE USED IN THE PRIMARY POSITION

The map advice of THIS CODE IS NOT TO BE USED IN THE PRIMARY POSITION is to be added when the SNOMED concept maps to an ICD-10 code that would only be used in a secondary position. This circumstance is different from the current advice THIS IS AN EXTERNAL CAUSE CODE FOR USE IN A SECONDARY POSITION.

Example

ICD-10 code: Z37.0, Single live birth

- THIS CODE MAY BE USED IN THE PRIMARY POSITION WHEN THE MANIFESTATION IS THE PRIMARY FOCUS OF CARE

Map advice THIS CODE MAY BE USED IN THE PRIMARY POSITION WHEN THE MANIFESTATION IS THE PRIMARY FOCUS OF CARE has been created for the 20150731 release. This advice is added to all asterisk codes. This advice applies regardless of whether the map is to a single code or part of a dagger and asterisk combination.

Only apply this map advice when the dagger and asterisk codes are the primary and secondary targets in the map.

Example
Diabetic
- neuropathy E14.4† G63.2*
- - polyneuropathy E14.4† G63.2*

Map to:
E14.4 Unspecified diabetes mellitus with neurological complications
G63.2 Diabetic polyneuropathy with the following advice:

- THIS CODE MAY BE USED IN THE PRIMARY POSITION WHEN THE MANIFESTATION IS THE PRIMARY FOCUS OF CARE.
- THIS IS AN EXTERNAL CAUSE CODE FOR USE IN A SECONDARY POSITION

An external cause code from Chapter XX is used with a code from another chapter to add to the detail captured by the diagnosis code by giving the reason for the condition, especially in situations where the diagnosis code specifies to “use additional external cause code.” The external cause code should always be sequenced AFTER the disease chapter code in a secondary position.

An exception to this rule is when a SNOMED Concept is described as an ‘Event’ and maps to only one external cause code. In this instance, the map advice THIS IS AN EXTERNAL CAUSE CODE FOR USE IN A SECONDARY POSITION should be added to identify both the need to record a diagnosis code and that sequencing rules apply.

- USE AS PRIMARY CODE ONLY IF SITE OF BURN UNSPECIFIED, OTHERWISE USE AS SUPPLEMENTARY CODE WITH CATEGORIES T20-T29 (BURNS)

Codes in categories T31 Burns classified according to extent of body surface involved, and T32 Corrosions classified according to extent of body surface involved, capture information about the percentage of body surface that has been burned or corroded, but codes in these categories should only be used as the main condition if the specific site of the burn is unknown. However, they can be used as an additional code to add more detail to a diagnosis.

The Mapping Specialist selects USE AS PRIMARY CODE ONLY IF SITE OF BURN UNSPECIFIED, OTHERWISE USE AS A SUPPLEMENTARY CODE WITH CATEGORIES T20-T29 (BURNS) map advice when mapping SNOMED concepts to codes in these categories.
The International Classification of Diseases requires that medical record coding include both dagger (aetiology) and asterisk (manifestation) codes together as a pair to describe certain conditions. In the alternative process of mapping from SNOMED CT, concepts may be void of the information required to capture both the dagger and asterisk code, but with a sufficient amount of information to enable the mapping of a lone asterisk code. Any advice has been created to address the issue of a single asterisk code map. THIS MAP REQUIRES A DAGGER CODE AS WELL AS AN ASTERISK CODE is applied to a map comprised of a single asterisk code. The advice tells the end user that another code is required without implying sequence order.

Example

202652006 |Inflammatory spondylopathy associated with another disorder (disorder)|
Map to:
M49.89 Spondylopathy in other diseases classified elsewhere Site unspecified, with the following advice:

- THIS CODE MAY BE USED IN THE PRIMARY POSITION WHEN THE MANIFESTATION IS THE PRIMARY FOCUS OF CARE, and
- THIS MAP REQUIRES A DAGGER CODE AS WELL AS AN ASTERISK CODE.

Types of map

SNOMED International provide two types of maps:

- **Simple map**: This links a single SNOMED CT concept to a single classification code to represent the clinical meaning of the concept.

Example 1

49436004|Atrial fibrillation (disorder) mapGroup 1| mapCategoryId =Properly classified", mapRule=TRUE, mapAdvice="ALWAYS I48", This is directly indexable in the ICD-10 Alpha index.

- **Complex map**: Links a single SNOMED CT concept to a combination of classification codes which when grouped together represents the meaning of the SNOMED CT concept. Source concepts that map to ICD-10 chapters with dagger
and asterisk conventions will be mapped to two target classification codes. The asterisk classification will be the second target record (mapGroup=2). The asterisk code will have applied the advice of THIS CODE MAY BE USED IN THE PRIMARY POSITION WHEN THE MANIFESTATION IS THE PRIMARY FOCUS OF CARE.

Example 2

111900000 | Pneumonia in aspergillosis (disorder) | maps to B44.1 Other pulmonary aspergillosis (dagger) and J17.2 Pneumonia in mycoses (asterisk):

- B44.1 “Other pulmonary aspergillosis” mapGroup 1, mapCategoryId="Properly classified", mapRule=TRUE, mapAdvice="ALWAYS B44.1", Map Target = “B44.1”
- J17.2 “Pneumonia in mycoses” mapGroup 2, mapCategoryId="Properly classified", mapRule=TRUE, mapAdvice= “ALWAYS J17.2” and “THIS CODE MAY BE USED IN THE PRIMARY POSITION WHEN THE MANIFESTATION IS THE PRIMARY FOCUS OF CARE” Map Target = “J17.2”

Definition of ambiguity

Cases for concern or question of ambiguity in the SNOMED CT source concept definition will include:

- Discrepancy between the Fully Specified Name and associated defining relationships; and
- Discordance between the SNOMED CT definition and the term synonyms.

Mapping Methodology

1. Evaluation of the SNOMED CT concept Fully Specified Name (FSN), defining relationships, (parents) and attributes to fully understand the semantic domain of the concept.
2. Location of the best semantic domain for the concept in ICD-10 using the Mapping Specialists coding knowledge along with any definition used to author the concept.
4. Consideration of ICD-10 Alphabetical Index essential modifiers and Tabular List exclusion notes to identify appropriate target maps.

5. The correct sequencing of codes in accordance with the rules and conventions of ICD-10.


7. Mapping rules have been incorporated into the mapping tool as a result of algorithmic mapping. This ensures consistency when adding map advice.

Review mechanism/Quality assurance

Validation of individual maps

1. A single map and review process of the SNOMED to ICD-10 is undertaken. One map specialist assigns the target map, and a second mapper reviews. There is feedback functionality within the mapping tool which allows for dialogue at the time of mapping/review.

2. If the mapper and reviewer cannot reach an agreement, a third mapper resolves any discrepancy.

3. The mapping tool and SNOMED authoring tools allow the mapper to view the authoring and history of the concept from receipt of a request to the construction of the concept. This vision of a concept and its attributes, body site involvement and the author concerned facilitates query raising as close to real-time authoring as possible.

4. Where there is a complete match, the mapped concept is sent to the pre-publication queue.

5. All map specialist/map review notes, rationales, research links and discussions between all parties are carried out within the tool providing an audit trail and evidence for decisions.

6. Discussion and feedback between mapper and reviewer do not accompany the map but is stored for internal review and future decision making.

Validation of Published maps

A recent comparison of maps distributed as part of the January 2018 SNOMED International Release and the April 2018 UKTC Release identified very few (less than 1%) discrepancy after considering use case and local mapping principles/guidelines.
Technical QA

The QA is comprised of:

- **Interactive Validation**: that is handled during the editing cycle. This QA is run each time a mapping specialist tries to save a concept and produces either errors (must be fixed) or warnings (take a second look). Rules are structural and content-based (e.g. record must end with TRUE rule, or target code must be assignable)

- **Workflow States Validation**: We have an admin QA to verify that only allowable workflow states exist. It is run regularly.

- **Database QA**: outside the application, we have a QA script to enforce and double-check conditions that are not bound by db level integrity constraint

- **Release QA**: At the end of the release cycle, a FINAL version of SNOMED will be obtained, and all changes reconciled (retired, out of scope, new in scope, changed in meaningful ways, etc.)

- **RF2 QA**: final review of RF2 files according to standing specification for QA developed at NLM

Mapping Resources

- SNOMED International Mapping Tool https://mapping.ihtsdotools.org/#/
- ICD-10 Volume 1 Tabular List (within the tool), Volume 2 Instruction book, Volume 3 Alphabetical Index
- WHO Online training
- SNOMED International Mapping Principles Handbook
- SNOMED International Terminology Browser
- Mapping Service Team Best Practice Document
- Approved websites (NICE, Orphanet, OMIM)

Release cycles and maintenance

- The maps are released twice a year in January and July as a part of the SNOMED CT International Release.
- The SNOMED CT International Release includes a mapping table from clinical concepts in SNOMED CT to codes listed in ICD-10.
- SNOMED CT to ICD-10 Maps are represented as members of a single Reference Set called the ICD-10 Complex Reference Set. This Reference Set contains all map member data.
- Agreed updated versions of ICD-10 released by the WHO Update and Revision Committee shall be subject to a revision of the MAP twice yearly and included in the next following SNOMED CT release.
- Incremental changes to the map shall be documented employing the SNOMED CT Enhanced Release Format.
- SNOMED International welcomes the opportunity to work with organizations or institutions interested in undertaking usage validation of the SNOMED CT to ICD-10 Map.

Competencies and skill set for mappers

- Knowledge of the source and target terminologies (Country specific qualification)
- Understand and explain the purpose of the map (SNOMED International e-learning programmes)
- Understand the way in which the computer system and people will use the map (user experience)
- Understand and be able to apply the structure, content and relationships for the source and target terminology
- Be able to apply the basic concepts of the SNOMED CT model and description logic. (SNOMED International Foundation e-learning course)
- Understand the processes to maintain and publish the map.

Strengths and limitations of mapping

- Fitness for purpose: The SNOMED CT maps directly link to ICD-10 using the coding rules and conventions as instructed. The map can be tailored for local use or country-specific guidelines.
- Subsets of SNOMED CT can be utilised.
- Information loss: As the source concept generally has a greater level of specificity and the map is to a clinical code with specific and specified/unspecified codes within a category, information loss is kept to a minimum. Instruction is also given by map advice to add specificity contained within the medical record where possible.
- Biases: SNOMED CT concepts are more specific than the ICD-10 code/s and offer more granularity. No biases identified.
- Resource requirements: SNOMED International Mapping Tool, Clinicians, Terminologists and Mappers.
- Method applicable to other use cases: Yes (members and affiliates of SNOMED International)
Appendix 4: NHS Digital (UK) SNOMED CT to ICD-10 Maps

Author: Hazel Brear, Principal Classifications Specialist, NHS Digital

Source terminology:
SNOMED CT

Target classification:
ICD-10 5th Edition

Purpose of this mapping\(^6\) (use cases supported)

The purpose of mapping is to produce a reliable, consistent and reproducible link, in one direction, from SNOMED CT to ICD-10 to support coding for morbidity purposes.

Because the classification is a vital component of national datasets and underpins the National Tariff payment system, it is crucial that organisations using an Electronic Patient Record (EPR) can efficiently derive ICD-10 codes semi-automatically from the clinical information recorded using SNOMED CT.\(^7\) The maps support the "record once, use many times" principle and reduce manual coding allowing classification experts to concentrate on the coding of more complex cases.

Mapping technique

This is a manual mapping created by human map specialists using defined resources related to SNOMED CT and ICD-10.

Documentation

We provide technical specification and implementation guidance and a map change report as a part of the release pack.

\(^6\) Mappings are created to both ICD-10 and OPCS-4, but for this case study, only ICD-10 is mentioned.

\(^7\) SNOMED CT is mandated as the single terminology for use in NHS systems in England for direct patient care management.
Publication and release

The maps are provided in Release Format 2 (RF2) as an artefact of the SNOMED CT UK Edition release in April and October each year via our distribution service Terminology Release and data Update and Distribution (TRUD).

The licenses for the individual components of the maps must be accepted before the download of the products.

Examples of usage:

The ICD-10 classification is a fundamental NHS Information Standard and an essential element of NHS patient activity data. The maps are designed to support Admitted Patient Care Commissioning Data Sets (CDS), Central Returns and other data sets. The ICD-10 codes derived from SNOMED CT concepts support population health care. For example:

- Population of Commissioning Datasets (CDS), e.g. Admitted Patient Care CDS
- Statistical reporting, e.g. Hospital Episode Statistics (HES), Secondary Uses Service (SUS)
- Generation of Healthcare Resource Groups (HRGs) and currencies which underpin the National Tariff System (reimbursement)
- Outcomes framework
- National policies and strategies provided by:
  - Public Health England (PHE)
  - NHS England (NHSE)
  - NHS Improvements (NHSI)
  - Department of Health and Social Care (DHSC)

What the maps are:

They provide a semi-automated selection of codes from ICD-10 to represent the multiplicity of circumstances encountered in the medical record. The maps from SNOMED CT to ICD-10 reflect the clinical meaning of a SNOMED CT concept, the rules and conventions of ICD-10, the UK National Clinical Coding Standards, the Three Dimensions of Coding Accuracy and approved Editorial Mapping Principles.
Scope

The scope of this map set is a defined subset of SNOMED CT hierarchies and concepts (the source) and all codes within the ICD-10 5th Edition classification (the target). Most concepts in the defined SNOMED CT subset are from the following hierarchies:

- Clinical finding (finding)
- Event (event)
- Situation with explicit context (situation)

**Note:** There are some exceptions with groups and individual concepts from other hierarchies included to ensure valid ICD-10 codes are reached (see Appendix 4.A).

Cardinality

The maps are a link from a single SNOMED CT concept to one or more ICD-10 target codes, which most closely represent the meaning of the source concept. Some of the maps offer alternatives (maps and targets) to represent information that is not available in the context of the concept but which may be available in the medical record. Consequently, there are four types of map with cardinality as follows:

- Map types 1 & 2 have a cardinality of \(1:1\)
- Map types 3 & 4 have a cardinality of \(1:many\)

(See “Types of map”)

Inputs:

- Individual SNOMED CT concept (source), ICD-10 codes (target)

---

8 Referred to in the technical specification document as "start points"

9 \(1:1\) is either one source code to one target code or one source code to more than one target codes, all of which are required to represent the meaning of the source code.

10 \(1:many\) is one source code to a choice of target codes (defaults and alternatives)
Output:

- Individual SNOMED CT concepts with associated ICD-10 codes and mapping attributes (see "Mapping attributes")
- Distributed as part of the NHS Digital Terminology released in Release Format 2 (RF2)

Relationships between inputs and outputs:

- The level of confidence is not currently recorded. A default target code or codes are the best semantic match.
- Alternative target codes may be offered to represent information unavailable in the context of the concept but which may be available to the end-user in the EPR.

Mapping attributes:

**referencedComponentId:** The ConceptID for the SNOMED CT concept being mapped

**mapBlock:** An integer identifying a number of groups of complex map records. There will be a minimum of one mapBlock and possibly multiple mapBlocks for each complex map. mapBlocks are specific to the NHS Digital maps.

**mapGroup:** An integer, grouping a set of complex map records from which one may be selected as a target code. Where a concept maps onto ‘n’ target codes, there will be ‘n’ groups, each containing one or more complex records.

**mapPriority:** Within a group, the mapPriority specifies the order in which complex map records should be checked. Only the first map record meeting the run time selection criteria will be taken as the target code within the group of alternate codes.

**mapRule:** A machine-readable rule (evaluating to either ‘true’ or ‘false’ at run time) that indicates whether this map record should be selected within its mapGroup. This data item is currently unused in the UK Edition data files, although it is populated in the UK Edition SNOMED CT Browser to indicate 'Alternative' or 'True' targets.

**mapAdvice:** Human-readable advice that may be employed by the software vendor to give an end-user advice on the selection of the appropriate target code or codes as well as prompting the addition of other codes. For example:

- ADDITIONAL CODE POSSIBLE
- ADDITIONAL CODE MANDATORY
- POSSIBLE REQUIREMENT FOR EXTERNAL CAUSE CODE
- MANDATORY REQUIREMENT FOR FIFTH DIGIT

*(see Appendix 4.B for a full list of current map advice)*

**mapTarget:** The target code in the scheme being mapped onto.

What the maps are not

The maps do not currently provide exhaustive resolution of all possible target codes. System supplier feedback suggests complex maps designed in this way are excluded when loading the maps into their [human readable] applications (encoders).

Types of map

We provide four different types of maps to represent the multiplicity of circumstances encountered in a medical record. In the following examples taken from the SNOMED CT UK Edition Browser, the **Map entries** are displayed as:

mapBlock/mapGroup/mapPriority e.g. 1/1/1 or 1/2/1 etc.

Map type 1

Links a single SNOMED CT concept to a single classification code to represent the clinical meaning of the concept.

Example

43339004 | Hypokalemia (disorder) (UK Edition Browser view)

<table>
<thead>
<tr>
<th>Map Entries</th>
<th>Rule</th>
<th>Advice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1/1</td>
<td>E87.6 Hypokalaemia</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

Map from a single SNOMED CT concept to a single ICD-10 code which represents the clinical meaning of the concept.

Map Type 2

Links a single SNOMED CT concept to a combination of classification codes which collectively represents the meaning of the SNOMED CT concept.
Example

7571000119109 | Thrombocytopenia co-occurrent and due to alcoholism (disorder)

<table>
<thead>
<tr>
<th>Map Entries</th>
<th>Rule</th>
<th>Advice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1/1</td>
<td>D69.5</td>
<td>TRUE</td>
</tr>
<tr>
<td>1/2/1</td>
<td>F10.2</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

Both codes are required to represent the meaning of the SNOMED CT concept
Sequencing is enforced by the mapGroup.

Map Type 3

Links a single SNOMED CT concept to a choice of classification codes (default(s) and alternative targets). Resolution of the choices involves a coding expert using the clinical detail found within the medical record (EPR), applying the rules, conventions and standards of the classifications and manually selecting the final classification code or codes from a list of alternative targets.

Example

49436004 | Atrial fibrillation (disorder) (UK Edition Browser view)

<table>
<thead>
<tr>
<th>Map Entries</th>
<th>Rule</th>
<th>Advice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1/1</td>
<td>I48.1</td>
<td>ALTERNATIVE</td>
</tr>
<tr>
<td>1/1/2</td>
<td>I48.2</td>
<td>ALTERNATIVE</td>
</tr>
<tr>
<td>1/1/3</td>
<td>I48.0</td>
<td>ALTERNATIVE</td>
</tr>
<tr>
<td>1/1/4</td>
<td>I48.9</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

With no other information than the meaning of this concept, the correct (default) code to select is I48.9. However, there are some essential modifiers for this condition, i.e. 'Chronic', 'Persistent' and 'Paroxysmal'. If the clinical coder sees any modifier detail in
the EPR, they will be able to select the correct code from the alternative target codes, i.e. I48.1, I48.2, I48.0.

Map type 4

Links a single SNOMED CT concept to a choice of classifications maps. Each choice of a map may contain a single combination or choice of target codes.

Example

10698009 | Herpes zoster iridocyclitis (disorder) (UK Edition Browser view)

<table>
<thead>
<tr>
<th>Map Entries</th>
<th>Rule</th>
<th>Advice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1/1</td>
<td>B02.3 Zoster ocular disease</td>
<td>TRUE</td>
</tr>
<tr>
<td>1/2/1</td>
<td>H22.0 Iridocyclitis in infectious and parasitic diseases classified elsewhere</td>
<td>TRUE</td>
</tr>
<tr>
<td>2/1/1</td>
<td>H22.0 Iridocyclitis in infectious and parasitic diseases classified elsewhere</td>
<td>TRUE</td>
</tr>
<tr>
<td>2/2/1</td>
<td>B02.3 Zoster ocular disease</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

A principle of ICD-10 is that a dagger code in a dagger and asterisk combination is the primary code and should be sequenced first. However, for morbidity coding, the dagger and asterisk sequence may be reversed when the manifestation of a disease is the primary focus of care.

In this example, there are two maps offered. The first (default) map has the dagger code B02.3 in the primary position with the additional asterisk code H22.0 in the secondary position (map outlined in green).

The second map reverses this sequencing and can be selected by the user if the iridocyclitis was the focus of care. (map outlined in orange).

Note: Map type 1 and 2 may be generated automatically within systems allowing the coding expert to devote time to the resolution of concepts with a map type 3 or 4 map.
Mapping Toolkit:

1. Customised tooling\(^{11}\) environment providing functionality for dual blind authoring (best practice, reduces bias) and integrated workflow (user guidance available).


4. Editorial Mapping Principles (within the tool) (see Appendix 4.C for examples).

5. UK Edition SNOMED CT Browser (includes published maps) (also linked from within the tool).

Mapping steps

1. Evaluation of the SNOMED CT concept Fully Specified Name (FSN), defining relationships, (parents) and attributes to fully understand the semantic domain of the concept.

2. Location of the best semantic domain for the concept in ICD-10 using the Four Step Coding Process.\(^{12}\)

3. Identification of a default target code or codes ensuring application of the rules and conventions of ICD-10 and the UK national clinical coding standards.

4. Consideration of ICD-10 Alphabetical Index essential modifiers and Tabular List exclusion notes to identify potential “alternative” targets.

---

\(^{11}\) Developed for national requirements from SNOMED International Open-Source Tool

\(^{12}\) Step 1 – Analyse medical terminology to determine the Lead term and Modifier(s)

Step 2 – Locate the Lead term in the Alphabetical Index (ICD-10 Volume 3)

Step 3 – Assign a tentative code using the Alphabetical Index (ICD-10 Volume 3)

Step 4 – Verify the code using the Tabular List
5. Correct sequencing in accordance with the rules and conventions of ICD-10 and in support of the Three Dimensions of Coding Accuracy.\textsuperscript{13}
6. Application of editorial mapping principles, including assignment of map advice.

\textbf{Review mechanism/Quality assurance}

1. Mapping of each concept is completed independently by two mapping specialists.
2. On completion of the mapping by both specialists, the tooling environment compares the output.
3. Where there is a complete match, the mapped concept is sent to the pre-publication queue.
4. Where there is a difference (map target(s), map block, map group, map priority, map rule, map advice), the concept is sent via workflow for map lead review and resolution.
5. All map specialist/map lead notes, rationales, research links and discussions between all parties are carried out within the tool, providing an audit trail and evidence for decisions.
6. Error keys are applied to true errors allowing retrospective analysis to inform training needs.
7. Consensus management meetings (Mapping QA) are convened to resolve complex issues, again providing notes, discussions, rationales and research links providing evidence for decisions.
8. Technical validation rules (must be fixed) and warnings (confirm correct) are applied automatically and in real-time within the tool and throughout the editing cycle.
9. Further QA rules and warnings are applied following the export of the maps from the tool as part of the terminology technical pre-release protocol.

\textsuperscript{13} Individual Codes, Totality of codes, Sequencing of codes – Codes must be organised in a sequence which is statistically consistent
Validation of published maps

Regular comparison of SNOMED International and UK Edition ICD-10 maps is carried out, and discrepancies fed back and discussed. The Latest validation reports less than 1% discrepancy, excluding differences due to national standards and principles.

Feedback or queries from users of the maps can be submitted through the Terminology and Classifications Delivery Service product support desk at information.standards@nhs.net

Release cycles and maintenance:

The maps are released twice a year (April and October) as a part of the SNOMED CT UK Edition Release. In-scope new content in the International release (January and July), as well as in-scope new content specific to the UK Edition, is mapped.

Drivers of maintenance of existing maps:

- Changes to SNOMED CT content
- Updates to ICD-10
- Department of Health and Social Care (DHSC) requirements
- National Casemix Office requirements
- NHS England requirements
- Updates to National Clinical Coding Standards
- Updates to Editorial Mapping Principles

Competencies and skill set for map specialists:

- Experienced Clinical Coder who has attained the UK National Clinical Coding Qualification (NCCQ)
- Has 3-5 years’ minimum post qualification experience working within an NHS Acute Trust environment, applying clinical coding skills
- Experienced NHS Digital Classification Specialist

Skill Level:

Trainee:

- Competencies as above
Intermediate:

- Competencies as above
- Two years mapping experience (essential)
- A good understanding of the structure of SNOMED CT (relationships, attributes, content model) (essential)
- A good understanding of the use of maps in secondary care (essential)
- Completed and passed the SNOMED CT Foundation eLearning Course (essential)
- Excellent understanding of the functionality of the mapping tool (essential)
- Excellent understanding of mapping editorial principles (essential)
- Excellent understanding of the different types of mapping roles (essential)
- Presentation to a national audience on the subject of mapping (desirable)
- Shadow Release Lead role (essential)
- Act as Release Lead for two release cycles (essential before progression to advanced)

Advanced

- Competencies as above
- 3-5 years mapping experience (essential)
- An excellent understanding of the structure of SNOMED CT (relationships, attributes, content model) (essential)
- An excellent understanding of the use of maps in secondary care (essential)
- Completed and passed the SNOMED CT Foundation eLearning Course (essential)
- Completed and passed the SNOMED International Content Development eLearning course (desirable)
- Confident as Release Lead (essential)

Annual Survey

NHS Digital seeks feedback from users of our products each year, and the annual survey contains questions about the quality of the maps and user satisfaction. Responses to the survey are collated and used to inform the annual service improvement plan.
Appendix 4.A

Many of the entities within ICD-10 Chapter XXI Factors influencing health status and contact with health services are viewed in the clinical terminology world as 'procedure' or 'situation' type concepts and, quite rightly, not suitable for inclusion as content in the clinical findings hierarchy. Therefore to ensure that the codes in this chapter are reached from SNOMED CT, we have 'pushed' several concepts from the SNOMED CT 'procedure', 'situations', 'context-dependent category', 'event' and 'qualifier value' hierarchies into the scope of ICD-10 and subsequently mapped them to ICD-10. These concepts have a map or excuse code to both ICD-10 and OPCS-4. This 'push-through' is not unprecedented as several concepts have carried dual mappings for many years.

Examples of concepts pushed into the scope of ICD-10:

<table>
<thead>
<tr>
<th>SNOMED CT Concept ID &amp; Fully Specified name</th>
<th>Mapped to ICD-10 target code</th>
</tr>
</thead>
<tbody>
<tr>
<td>171126009</td>
<td>Tuberculosis screening (procedure)</td>
</tr>
<tr>
<td>79841006</td>
<td>Genetic counselling (procedure)</td>
</tr>
<tr>
<td>105480006</td>
<td>Refusal of treatment by patient (situation)</td>
</tr>
<tr>
<td>735934008</td>
<td>History of poor personal hygiene (situation)</td>
</tr>
<tr>
<td>182918009</td>
<td>Repeated prescription (context-dependent category)</td>
</tr>
<tr>
<td>269772004</td>
<td>Contaminant given to patient (event)</td>
</tr>
<tr>
<td>416085006</td>
<td>Nosocomial transmission (qualifier value)</td>
</tr>
</tbody>
</table>
Appendix 4.B

The following map advices are available within the mapping tool and found in the mapping tables:

<table>
<thead>
<tr>
<th>ICD-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADDITIONAL CODE MANDATORY</td>
</tr>
<tr>
<td>ADDITIONAL CODE POSSIBLE</td>
</tr>
<tr>
<td>ADDITIONAL CODE POSSIBLE FOR OUTCOME OF DELIVERY</td>
</tr>
<tr>
<td>ADDITIONAL CODE POSSIBLE TO FULLY DESCRIBE DISEASE OR CONDITION</td>
</tr>
<tr>
<td>ADDITIONAL CODE POSSIBLE TO IDENTIFY CAUSATIVE AGENT</td>
</tr>
<tr>
<td>ADDITIONAL CODE POSSIBLE TO IDENTIFY PRESENCE OF HYPERTENSION</td>
</tr>
<tr>
<td>ADDITIONAL CODE POSSIBLE TO IDENTIFY PRESENCE OF RENAL FAILURE</td>
</tr>
<tr>
<td>ADDITIONAL CODE POSSIBLE TO IDENTIFY SEPTIC SHOCK</td>
</tr>
<tr>
<td>ADDITIONAL DAGGER CODE MANDATORY</td>
</tr>
<tr>
<td>ADDITIONAL EXTERNAL CAUSE CODE POSSIBLE</td>
</tr>
<tr>
<td>CODE MUST NEVER BE USED IN A PRIMARY POSITION</td>
</tr>
<tr>
<td>CODE MUST BE USED IN A PRIMARY POSITION</td>
</tr>
<tr>
<td>FIFTH CHARACTER MANDATORY</td>
</tr>
<tr>
<td>FIFTH CHARACTER POSSIBLE</td>
</tr>
<tr>
<td>HIGH LEVEL CONCEPT NOT CLASSIFIABLE</td>
</tr>
<tr>
<td>PERMISSIBLE TO USE THIS ASTERISK CODE IN A PRIMARY POSITION</td>
</tr>
<tr>
<td>THIS IS AN EXTERNAL CAUSE CODE FOR USE IN A SECONDARY POSITION</td>
</tr>
</tbody>
</table>
Appendix 4.C

Editorial Mapping principles drive consistency and reproducibility. For example:

**MPD 010: External cause codes**

<table>
<thead>
<tr>
<th>Status</th>
<th>Approved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implementation date</td>
<td>01-Apr-2017</td>
</tr>
</tbody>
</table>

Supplementary Chapter XX external cause codes in the range V01-Y36 are not included in the map (with the exception of poisoning type concepts) even when an indication or statement of cause is included in the context of the concept. The map advice ‘ADDITIONAL EXTERNAL CAUSE CODE POSSIBLE’ is added to the map.

This is to comply with National Clinical Coding Standard DChS.XX.1: External causes, which states: “Codes in categories V01-Y36 must only be assigned on the first consultant episode in which the condition is recorded. Any subsequent episode where the same condition is being treated does not require the external cause code from V01-Y36.”

**Example**

426246007 | Penetrating injury caused by knife (disorder)

In this case, even though the external cause (knife wound) has been stated in the FSN, the external cause code is not included in the map, but only the map advice “ADDITIONAL EXTERNAL CAUSE CODE POSSIBLE”.

![Map Entry](image-url)
Appendix 5: Australian Classification of Health Interventions (ACHI) to International Classification of Health Interventions (ICHI)

Author: Anupama Ginige, Western Sydney University, Australia

Source Terminology

Australia Classification of Health Interventions (ACHI)

Target Terminology

International Classification of Health Interventions (ICHI)

Purpose of this Mapping

- Support longitudinal morbidity statistics
- Interoperability for International epidemiological reporting

Example of Usage

- Support the comparison of data coded with ACHI with data coded with ICHI in the future.

What the Maps Are

The maps will be a linkage from a source national classification, ACHI, to a target international classification, ICHI, in one direction. The intervention description is matched to one or more ICHI stem code(s) and any corresponding extension code(s). Coding rules of inclusion terms, include notes, code also, and exclude notes are used as necessary.

Inputs

Source ACHI codes and its full description and its parent hierarchical and target ICHI stem codes and extensions. A complete set of ACHI codes and a subset of ICHI codes are the scope for mapping. Each ICHI axis has its stem code, title, inclusions and exclusions. ICHI interventions on the environment are excluded from the maps.
Outputs

ACHI codes with descriptions and corresponding ICHI codes with descriptions and comments.

Types of Map

There are two types of maps representing the types of health interventions performed:

1. Simple map – A single ACHI code maps to a single ICHI stem code plus relevant extension codes to represent the full clinical description of the intervention
2. Complex map – A single ACHI code maps to multiple ICHI stem codes or stem codes plus extension code, or multiple ACHI codes to one or more ICHI stem codes plus relevant extension codes

The Cardinality of a simple map is one-to-one. In contrast, a complex map is one-to-many, many-to-one or many-to-many, where the union of concepts represented by the "many" codes (and their extensions in the case of ICHI) is equivalent semantically to the concept represented on the other side of the map.

Mapping Toolkit

- ACHI 10th Edition
- ICHI 2018 Beta version
- Draft ICHI guidelines
- Indexed ICHI codes in ElasticSearch
- Medical Subject Headings (MeSH)

Mapping Methodology

In mapping ACHI to ICHI, mainly an algorithmic approach is used with input from classification specialists for validation and verification purposes. The mapping methodology used in this use case is depicted in Figure 1.

Figure 1 Process Flow
ACHI codes are arranged into anatomical and procedural axes. In the four-step process in Figure 1, a rule-based approach is initially used to map the ACHI axes into the three ICHI axes target, action, and means. This step deploys the natural language processing (NLP) system cTAKES™ [1] to divide the full ACHI descriptions into noun and verb chunks and to recognise MeSH (Medical Subject Headings) headings to provide a hierarchical context to the nouns and verbs [2]. Preliminary tests suggest that about 10% of the codes will be mapped in the first step based on their lexical similarity. These mapped codes are presented to a classification specialist for verification (Step 2 of the process) through a web interface, as shown in Figure 2. The verified (5%) and incorrect (5%) maps are used to create the initial training data that feeds into the subsequent machine learning step.

In Step 3, vectorisation of ACHI descriptions using the continuous bag of words (CBOW) [3] followed by the use of Skip-Gram algorithm [4] in combination with the use of Apache cTAKES™[1] to identify noun and verb in ACHI descriptions allow us to attempt mapping the 90% of ACHI codes that was not mapped in Step 1. Based on the pilot testing results, it is anticipated that around 80% of ACHI codes will be mapped in this step, and 20% will be unmapped. The process's final step involves classification specialists validating all automatically mapped files and manually mapping the 20% that was not mapped.

Figure 2 Example of ACHI to possible ICHI mappings
Mapping is carried out on a chapter-by-chapter basis. Therefore, all these mapping results of Step 2 and Step 4 are fed back into the machine learning Step 3 as training data, enhancing the accuracy of mappings of other chapters and future mappings.

Quality Assurance/Validation

Experienced clinical classification specialist will validate and verify maps in Step 2 and Step 4 stages of the process. The validated chapter is used as input for the following machine learning algorithm cycle.

Drivers for Maintenance of Existing Maps

- ICHI update cycle
- ACHI revision cycle

Competencies and Skill Set for Map Specialists

- A Highly experienced clinical classification specialist will carry out the verification and validation of the maps.
- Experts in machine learning, with some background knowledge in clinical classification, will develop the computerised algorithms.

References

1. cTAKES™, A. Apache cTAKES™ is a natural language processing system for extraction of information from electronic medical record clinical free text.; Available from: http://ctakes.apache.org/whycTAKES.html.


Appendix 6: International Classification of Diseases, Ninth Revision, Clinical Modification Volume 3 (ICD-9-CM Vol 3) to International Classification of Health Interventions (ICHI)

Author: Megan Cumerlato, Secretariat, ICHI Development Project

Source terminology:
International Classification of Diseases, Ninth Revision, Clinical Modification Volume 3 (ICD-9-CM Volume 3)

Target terminology:
International Classification of Health Interventions (ICHI) Beta 2018 version.

Purpose of this mapping:
The purpose is:

- Surgical procedure (shortlist) within ICD-9-CM Volume 3 mapping to ICHI Beta 2018.
- Investigate the feasibility and comprehensiveness of using ICHI for intervention coding purposes.

Examples of usage:
The mapping supports:

- Lay the foundation for possible comparison of clinical data already coded with ICD-9-CM Volume 3, with ICHI coded data in the future.

What the maps are:
The maps are a link between two international classification systems ICD-9-CM Volume 3 and ICHI Beta 2018. ICD-9-CM Volume 3 is the source classification, and ICHI is the target classification. The clinical content of the ICD-9-CM Volume 3 codes is matched to the clinical content of ICHI stem code(s) and extension code(s). Codes are also mapped by matching inclusion terms.
Inputs:

Outputs:
ICD-9-CM Volume 3 codes with label and description of the codes and their associated ICHI codes and extension codes with their description.

Relationship between inputs and outputs:
The match is designated as:
- Source code is equivalent to target code (E)
- Source code has a broader meaning than target code (B)
- Source code has a narrower meaning than the target code (N)
- There is no match at all (X)

The cardinality of the map is designated as:
- 1:1 with or without extension code(s)
- 1:many with or without extension code(s)
- many:1 with or without extension code(s)
- many:many with or without extension code(s)

In each cardinality designation, the source may be equivalent, broader, or narrower than the target.

Mapping attributes:

<table>
<thead>
<tr>
<th>Field</th>
<th>Data Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Code ID</td>
<td>Label</td>
<td>51.23</td>
</tr>
<tr>
<td>Source Code Description</td>
<td>String</td>
<td>Laparoscopic cholecystectomy</td>
</tr>
<tr>
<td>Target Code ID</td>
<td>String</td>
<td>KCF JK AB</td>
</tr>
<tr>
<td>Target Code Description</td>
<td>String</td>
<td>Laparoscopic cholecystectomy</td>
</tr>
</tbody>
</table>
Symbol of combination | Operator | A forward-slash (/) separating stem codes or an ampersand (&) following a stem code and preceding an extension code, or between multiple extension codes in a complex map
---|---|---
Relation | String | Identifies the level of equivalence of the map: equivalent, broader, narrower or no equivalence
Cardinality | String | Identifies complexity of map: 1:1, 1:many, many:1, or many:many

What the maps are not:
The maps provide alternative target codes, but not the semantic equivalency of the maps.

Types of maps:
There are two types of maps representing the types of health interventions performed on or for a person:

1. Simple map—links a single ICD-9-CM Volume 3 code to a single ICHI code with or without extension code(s) to represent the whole clinical meaning of the intervention
2. Complex map—links an ICD-9-CM Volume 3 code to multiple ICHI stem codes with or without extension code(s) or multiple ICD-9-CM Volume 3 codes to one or more ICHI stem codes with or without extension code(s) to represent the whole clinical meaning of the intervention.

Sample Map:

1. **1:1 map**: Links single ICD-9-CM Volume 3 source code to single ICHI target code.

Table 1: 1:1 map without extension code(s)

<p>| Source: ICD-9-CM Volume 3 | Target: ICHI Beta 2018 |</p>
<table>
<thead>
<tr>
<th>Source Code</th>
<th>Source Code Description</th>
<th>Target Code</th>
<th>Target Code Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.1</td>
<td>Intracapsular extraction of lens</td>
<td>BBFJK.AA</td>
<td>Intracapsular extraction of lens</td>
</tr>
<tr>
<td>13.2</td>
<td>Extracapsular extraction of lens by linear extraction technique</td>
<td>BBFJK.AH</td>
<td>Extracapsular extraction of lens</td>
</tr>
<tr>
<td>36.1</td>
<td>Bypass anastomosis for heart revascularisation</td>
<td>HIA.LI.AA</td>
<td>Coronary artery bypass, open approach (Direct revascularisation cardiac with catheter stent, prosthesis, or vein graft; Direct revascularisation coronary with catheter stent, prosthesis, or vein graft; Direct revascularisation heart muscle with catheter stent, prosthesis, or vein graft; Direct revascularisation myocardial with catheter stent, prosthesis, or vein graft)</td>
</tr>
</tbody>
</table>

Table 2: 1:1 map with extension code(s)

<table>
<thead>
<tr>
<th>Source: ICD-9-CM Volume 3</th>
<th>Target: ICHI Beta 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Code</td>
<td>Source Code Description</td>
</tr>
</tbody>
</table>

113
<table>
<thead>
<tr>
<th>Source Code</th>
<th>Source Code Description</th>
<th>Target Code</th>
<th>Target Code Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>81.51</td>
<td>Total hip replacement</td>
<td>MLJ.ML.AA&amp;XDE7</td>
<td>Hip replacement, total</td>
</tr>
<tr>
<td>53.0</td>
<td>Unilateral repair of inguinal hernia</td>
<td>PAM.MK.AA&amp;XCA9</td>
<td>Repair of inguinal hernia, unilateral</td>
</tr>
</tbody>
</table>

2. **1: Many map**: Links single ICD-9-CM Volume 3 source code to multiple ICHI codes. The tables below illustrate 1: Many maps with and without extension code(s).

### Table 3: 1:many map without extension code(s)

<table>
<thead>
<tr>
<th>Source Code</th>
<th>Source Code Description</th>
<th>Target Code</th>
<th>Target Code Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>28.3</td>
<td>Tonsillectomy with adenoidectomy</td>
<td>DAA.JK.AC / DAB.JK.AC</td>
<td>Tonsillectomy with adenoidectomy</td>
</tr>
<tr>
<td>20.96</td>
<td>Implantation or replacement of cochlear prosthetic device, not otherwise specified</td>
<td>CCB.DN.AC / CCB.JD.AC</td>
<td>Implantation of cochlear prosthetic device/Removal of cochlear prosthetic device</td>
</tr>
</tbody>
</table>

### Table 4: 1:many map with extension code(s)

<table>
<thead>
<tr>
<th>Source Code</th>
<th>Source Code Description</th>
<th>Target Code and extension codes</th>
<th>Target Code Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.97</td>
<td>Implantation or replacement of cochlear prosthetic device, single channel</td>
<td>CCB.DN.AC / CCB.JD.AC&amp;XT03.01</td>
<td>Implantation of cochlear prosthetic device/Removal of cochlear prosthetic device</td>
</tr>
</tbody>
</table>
3. **many : 1 map:** Links many ICD-9-CM Volume 3 source codes to a single ICHI code. Given the tables below gives an example of **many : 1 map** without extension code(s).

**Table 5: many : 1 map without extension code(s)**

<table>
<thead>
<tr>
<th>Source Code</th>
<th>Source Code Description</th>
<th>Target Code</th>
<th>Target Code Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>47.0, 47.1</td>
<td>Appendectomy (47.0)</td>
<td>KBOJK.AA</td>
<td>Appendicectomy (Other incidental appendectomy)</td>
</tr>
<tr>
<td></td>
<td>Incidental appendectomy (47.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47.01, 47.11</td>
<td>Laparoscopic appendectomy (47.01)</td>
<td>KBOJK.AB</td>
<td>Laparoscopic appendicectomy (Laparoscopic incidental appendectomy)</td>
</tr>
<tr>
<td></td>
<td>Laparoscopic incidental appendectomy (47.11)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. **many : many map:** Links many ICD-9-CM Volume 3 codes to many ICHI codes. Given the tables below gives an example of **many : many maps** without extension code(s).

**Table 6: many : many map without extension code(s)**

<table>
<thead>
<tr>
<th>Source Code</th>
<th>Source Code Description</th>
<th>Target Code</th>
<th>Target Code Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Description</td>
<td>HIA.LG.AF/HIA.LH.AF</td>
<td>Description</td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td>---------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>36.01, 36.02, 36.05</td>
<td>Single vessel percutaneous transluminal coronary angioplasty [PTCA] or coronary atherectomy without mention of thrombolytic agent (36.01) Single vessel percutaneous transluminal coronary angioplasty [PTCA] or coronary atherectomy with mention of thrombolytic agent (36.02) Multiple vessel percutaneous transluminal coronary angioplasty [PTCA] or coronary atherectomy performed during the same operation, with or without mention of thrombolytic agent (36.05)</td>
<td>HIA.LG.AF/HIA.LH.AF</td>
<td>Percutaneous transluminal coronary angioplasty or coronary atherectomy, with insertion of stent</td>
</tr>
</tbody>
</table>

Mapping toolkit:
- ICHI beta version 2018
- Draft ICHI guidelines

Mapping methodology:
The mapping between these classification systems was done manually by a clinical classification specialist. Steps involved in this mapping were:

1. Select an ICD-9-CM code and check the code label and description.
2. Locate the best semantic match(es) in ICHI, reviewing the target, action, means axes and using the Draft ICHI Guidelines for users and search function.

3. If no semantic match is available, leave the axis code blank.

4. If multiple possible semantic matches are available, select the closest match. This would mean selecting either one or more codes from the target classification that gives the closest semantic similarity.

5. Define the cardinality of the source to target map (1:1 or 1:many).

Quality assurance/validation:

❖ All mappings are done manually by classification specialists and verified by one other classification specialist.

Release cycles and maintenance:

The maps are created and released only once, and ICHI changes may require updating the mapping files.

Drivers for maintenance of existing maps:

ICHIC Release Cycles

Competencies and skill set for map specialists:

Experienced Clinical classification specialists

Skill level within the team:

● Advanced
Appendix 7: National Intervention Classification to the International Classification of Health Interventions

Authors: Karen Coghlan, Cassandra Linton, Sharon Baker, Canadian Institute for Health Information

Source terminology:

Canadian Classification of Health Interventions (CCI)

The Canadian Classification of Health Interventions, referred to as CCI, is a multi-axial classification of health-related interventions, developed and maintained by the Canadian Institute for Health Information (CIHI). CCI’s tabular listing provides comprehensive coverage of diagnostic, therapeutic, and other associated healthcare interventions, excluding laboratory and pathology procedures upon specimens and compounding and manufacturing of health-related devices and products. It is designed to be provider and location-neutral to be used across the continuum of healthcare settings in Canada.

Target terminology:

International Classification of Health Interventions (ICHI) Beta 2018

Purpose of this mapping (use cases supported):

There are two use cases for this set of maps:

- conversion from a national to an international classification to support longitudinal morbidity statistics
- interoperability for international epidemiological reporting (e.g., OECD surgical procedures).

The first use case involves mapping the entire set of CCI codes to ICHI codes for a gap analysis to determine if ICHI is fit for use in Canada. The second use case involves mapping a subset of CCI codes, as identified by the Organization for Economic Cooperation and Development (OECD) for sentinel surgical procedures, to ICHI to determine whether the maps can be used for international information exchange.
Examples of usage:

There is 100% coverage of hospital inpatient and day surgery major interventions and partial coverage of emergency room and ambulatory care clinic interventions in Canada using CCI codes. The maps must support:

➢ Hospital statistics and regional comparisons: inpatient, emergency and ambulatory care
➢ Case mix grouping for case costing and hospital reimbursement
➢ Health indicators as produced by CIHI and other Canadian organisations
➢ Joint replacement and transplant registries
➢ Secondary use for administrative data requests by researchers, graduate students, insurance, pharmaceutical and medical device companies, media and health ministries

What the maps are:

The maps are a linkage from a source (legacy) national classification, CCI, to a target international classification, ICHI, in one direction. The clinical content of the CCI code and any mandatory attribute concerning the status, location or extent of the intervention is matched to the clinical content of one or more ICHI stem code(s) and any corresponding extension code(s). Coding rules of inclusion and exclusion are reflected in the map as necessary to explicate context.

Inputs:

Source CCI codes and mandatory (status, location, mode of delivery, extent) attributes and target ICHI stem codes and extensions. A complete set of CCI codes and a subset of ICHI codes are in scope for mapping. ICHI Interventions on the environment are excluded from the maps.

Outputs:

CCI codes and mandatory attributes with long descriptions and associated ICHI codes and extensions with long descriptions and mapping attributes.

Relationship between inputs and outputs:

The semantic match is designated as:
- source code is equivalent (lexical or synonymous) to target code (E)
- source code is broader in meaning than the target code (B)
- source code is narrower in meaning than the target code (N)
- there is no semantic match at all (X)

The map’s cardinality is designated as 1:1 (simple) or 1:many or many:1 or many:many (complex). The Totality and sequence of complex maps are dependent on the rules and conventions of both classifications, which affect the designation of the semantic match. The Level of confidence is not rated because it is dependent on the use of the map, which is variable.

Mapping attributes:

<table>
<thead>
<tr>
<th>Field</th>
<th>Data Type</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Map Group</td>
<td>Integer</td>
<td>Identifies the number of codes and mandatory attributes that comprise the source concept.</td>
</tr>
<tr>
<td>Source Code ID</td>
<td>String</td>
<td>Alphanumeric CCI code unpunctuated and mandatory code attribute(s)</td>
</tr>
<tr>
<td>Source Code Description</td>
<td>String</td>
<td>CCI long description of code and mandatory attribute(s)</td>
</tr>
<tr>
<td>Symbol of Combination</td>
<td>Operator</td>
<td>An ampersand (&amp;) symbol following a source code and preceding a mandatory attribute code or between multiple mandatory attribute codes in a complex map</td>
</tr>
<tr>
<td>Map Priority</td>
<td>Integer</td>
<td>Specification of the order in which complex maps are checked with first source code mapping to first target code</td>
</tr>
<tr>
<td>Map Rule</td>
<td>String</td>
<td>A machine-readable rule evaluating code description equivalence (true/false) for promotion of candidate map</td>
</tr>
<tr>
<td>Map Advice</td>
<td>String</td>
<td>Human-readable advice applied by the mapper and employed by vendor software for end-user advice on selection of additional target codes (e.g. code also)</td>
</tr>
<tr>
<td>Target Code ID</td>
<td>String</td>
<td>Alphanumeric ICHI stem code punctuated and extension code(s)</td>
</tr>
<tr>
<td>---------------</td>
<td>--------</td>
<td>-----------------------------------------------------------</td>
</tr>
<tr>
<td>Target Code Description</td>
<td>String</td>
<td>ICHI long description of stem and extension code(s)</td>
</tr>
<tr>
<td>Symbol of Combination</td>
<td>Operator</td>
<td>A forward-slash (/) separating stem codes or an ampersand (&amp;) following a stem code and preceding an extension code, or between multiple extension codes in a complex map</td>
</tr>
<tr>
<td>Target Map Group</td>
<td>Integer</td>
<td>Identifies the number of stem codes and extension codes that comprise the target concept</td>
</tr>
<tr>
<td>Map Relation</td>
<td>String</td>
<td>Identifies the level of equivalence of the map: equivalent, broader, narrower or no equivalence</td>
</tr>
<tr>
<td>Map Cardinality</td>
<td>String</td>
<td>Identifies complexity of map: 1:1, 1:many, many:1, or many:many</td>
</tr>
</tbody>
</table>

What the maps are not:

The maps do not provide alternative target codes, only the best semantic matches. Not all target codes in ICHI are selected for this mapping table.

Types of code map:

Unlike other national intervention classifications to ICHI mappings described in this white paper, where a source expression consisting of an intervention code and one or more attributes is considered a single mapping entity, a CCI code plus an attribute code is considered to have the cardinality of “many”. Similarly, an ICHI stem code plus one or more extension codes is considered to have the cardinality of “many” for this mapping.

Based on this definition of cardinality, there are two types of maps:

1. **Simple code map**—links a single CCI code to a single ICHI code to represent the intervention’s full clinical description. The source code may be equivalent, broader or narrower than the target code.
<table>
<thead>
<tr>
<th>Source Code and mandatory attributes</th>
<th>Source Code Description</th>
<th>Target Code and extension codes</th>
<th>Target Code Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.ZZ.02.ZU</td>
<td>Assessment (examination), total body for physiological function only (primarily)</td>
<td>SH1.AA.ZZ</td>
<td>Assessment of mobility</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cardinality of Map</th>
<th>Relationship of Source Code to Target Code (Equivalence)</th>
<th>Human-readable Map Advice:</th>
</tr>
</thead>
<tbody>
<tr>
<td>One to one</td>
<td>Broader</td>
<td>True</td>
</tr>
</tbody>
</table>

Discussion:

Rehabilitation assessment is more broadly defined in CCI to refer to all physiological functions. As there are no special reporting requirements using this code, a simple map is sufficient.

2. **Complex code map**—links a single CCI code (plus one or more attributes) to multiple ICHI stem codes or one stem code plus an extension code to collectively represent the intervention's full clinical description.

A simple map's cardinality is 1:1, and a complex map is 1:many, many:1 or many:many.
<table>
<thead>
<tr>
<th>Source: CCI Version 2018</th>
<th>Target: ICHI Beta Version 2018 (or later)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Code and mandatory attributes</td>
<td>Source Code Description</td>
</tr>
<tr>
<td>1VA53LAPM</td>
<td>Implantation of device, hip joint open approach (direct lateral, posterolateral, posterior, transgluteal) single component prosthetic device [femoral]</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Revision</td>
</tr>
<tr>
<td>L</td>
<td>Left</td>
</tr>
<tr>
<td>FH</td>
<td>Modular ball (with or without modular neck)</td>
</tr>
</tbody>
</table>

**Cardinality of Map** | **Relationship of Source Code to Target Code (Equivalence)** | **Human-readable Map Advice:**
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>many:many</td>
<td>Narrower</td>
<td>Requires post-coordination</td>
</tr>
</tbody>
</table>

**Discussion:**

This map meets the need of Casemix grouping: 317 Revised hip replacement but does not meet the need of the Canadian Joint Replacement Registry for the type of prosthesis: single component, modular ball. There is no extension code available for this level of detail.
Preliminary field testing results from mapping 100 CCI codes to ICHI that represent the highest usage in hospital inpatient and ambulatory care in Canada yielded the following:

Relationship of source to target maps:
Equivalent 58%; Narrower 39%; Broader 3%; None 0%

Cardinality of maps:
1:1 70%; 1:many 18%; many:1 7%; many:many 5%

Mapping toolkit:
- CCI version 2021 within Classification Information Management System (CIMS)
- ICHI beta version 2018 (or later as available)
- Draft ICHI guidelines
- Mapping terminology tool (TBD) containing access to code sets and integrated workflow
- Mapping principles contained in wiki or mapping terminology tool (TBD)

Mapping methodology:
1. Select a CCI code and review the code components: anatomy/function/stage of pregnancy, intervention, approach/technique, device/agent used, tissue used. Also, check for any mandatory code attributes: status, mode of delivery, location, extent.
2. Locate the best semantic match(es) in ICHI, reviewing the target, action, means axes and using the draft ICHI, guidelines and search function.
3. If no semantic match is available, leave the target code blank.
4. If multiple possible semantic matches are available, select the closest lexical or synonymous match and flag with a comment for later review.
5. Sequence codes (for complex maps) according to rules identified in mapping principles.
6. Identify the relationship and cardinality of the source to target map.
7. Add any human-readable mapping advice for end-users that will be applicable for this map (e.g. code also instructions).
Quality assurance/validation:

- Simple (1:1) maps with equivalent relationship, either lexical or synonymous matching of source and target code titles, are automatically mapped within terminology mapping tool and verified by one mapping specialist.
- Complex (1:many, many:1, many:many) maps are completed manually by two mapping specialists in double-blinded, parallel fashion.
  - On completion of the mapping by both specialists, the terminology mapping tool compares the output.
  - If both maps match exactly, the map is sent to the pre-publication queue.
  - If there is a discrepancy of any type, the map is sent via workflow for map lead review and resolution.
  - All map notes, rationales, research links and team discussions are documented within the terminology mapping tool to provide evidence for decisions and a permanent audit trail for historical reference.
  - Error keys are applied to true errors to allow for retrospective analysis and inform training needs.
  - Consensus management team meetings are convened to resolve complex issues and update mapping principles as necessary.
- Technical validation rules and warnings are applied automatically and in real time within the terminology mapping tool and throughout the editing cycle.
- Further QA rules and warnings are applied following the maps' export from the terminology mapping tool as part of the terminology technical pre-release protocol.

Release cycles and maintenance:

The maps are released on a 3-year cycle (April) as part of the CCI update cycle: next release is 2021.

Drivers for maintenance of existing maps:

- ICHI update cycle
- Updates to mapping principles
- Casemix requirements
- Indicator requirements
- Registry requirements
● International research (e.g. OECD) requirements

Competencies and skill set for map specialists:

Experienced Health Information Professional who has attained certification by the Canadian Health Information Management Association (CHIMA) or Association Québécoise des Archivistes Médicale (AQAM).

- With 3-5 years, minimum post-qualification experience coding both inpatient and ambulatory care cases in a Canadian hospital.
- Experienced CIHI Classification Specialist
- Knowledge of both the source and target classifications and rules for use

Skill level within the team:

- Trainee (no mapping experience)
- Intermediate (at least one mapping project completed)
- Advanced (multiple mapping projects and types of mapping completed)
Appendix 8: Swedish Classification of Health Care Interventions to the International Classification of Health Interventions

Author: Ann-Helene Almborg, National Board of Health and Welfare, Sweden

Source terminology:
Swedish Classification of Health Care Interventions (KVÅ)

Target Terminology:
International Classification of Health Interventions (ICHI)

Purpose of this Mapping (use case supported):
The National Board of Health and Welfare (NBHW) has performed a pilot study regarding mapping subsets of the Swedish Classification of Health Care Interventions (KVÅ) (Source classification) to ICHI (Target classification) during 2018. In Sweden, the NBHW is responsible for maintaining and updating KVÅ. This mapping has been a part of the informal ICHI test trials.

There are two use cases for this pilot study:
1. to check how two sets of interventions in KVÅ match ICHI
2. to test and evaluate the mapping rules

The use case involved mapping 50 surgical orthopaedic interventions and 50 functioning interventions (both diagnostic and therapeutic) to understand KVÅ and ICHI’s relation (only this direction). The pilot study results will form the basis for further work to develop a plan to evaluate the consequences of converting from using KVÅ to using ICHI.

KVÅ is professionally neutral and consists of two parts:
1. Surgical interventions (about 6 000 interventions) based on Nomesco surgical procedure classification (NCSP), distributed into 19 chapters of interventions and one chapter of additional codes
2. Non-surgical interventions (about 3 600 interventions), including medical and functioning interventions, distributed into seven chapters of interventions, one
chapter of additional information and one chapter of standardised care process. Two chapters are based on two axes: Target (related to ICF) and Actions.

Examples of usage:

Today KVÅ is used for reporting to the national patient register interventions performed during inpatient encounters and visits to physicians in open specialised care. KVÅ is also used in electronic health record (EHR), which support information sharing among different caregivers, local follow-up, management and reporting to quality registers.

What the maps are:

Two subsets of interventions (50 orthopaedic surgical interventions and 50 functioning interventions) in KVÅ (source classification) were mapped the ICHI (target classification) in one direction. Codes of the source interventions were mapped to ICHI stem code(s) and extension codes if needed to cover the content in the source interventions. The mapping rules for ICHI were used, and cardinality was also added. The mapping was performed by two mappers manually.

Inputs:
The subsets of 100 KVÅ interventions in Swedish (codes, titles, inclusion, and exclusion) and the functioning interventions also have descriptions. The functioning interventions are based on two axes (Target [related to ICF] and Actions). KVÅ is included in the NCSP+ system (NOMESCO Classification of Surgical Procedures), and codes and titles (in English) were used. The whole ICHI, including its axes, stem codes and extension codes, was used in the mapping.

Outputs:

KVÅ interventions (codes, titles and descriptions incl NSCP+ codes) mapped to the axes of ICHI and ICHI stem codes with or without extension codes and the mapping results.

Relationship between source and target interventions.
The degree of equivalence of the concepts (not lexical) according to ICHI mapping rules was used such as:

- source intervention is equivalent (concept) to target intervention (E)
- source intervention is broader in meaning than the target intervention (B)
- source intervention is narrower in meaning than the target intervention (N)
- there is no semantic match at all (X), or there is no target intervention to be used

Mapping also includes mapping to each axis in ICHI and to value degree of equivalence. The map’s cardinality is designated as $1:1$ or $1:\text{many}$ ICHI stem codes (with or without extension codes).

Types of map:

There are two types of map representing the types of health interventions performed on or for a person:

1. Simple map—links a single KVÅ code to a single ICHI stem code (with or without extension code(s)) to represent, as much as possible, the full clinical description of the intervention.
2. Complex map—links a single KVÅ code to multiple ICHI stem codes (with or without extension code(s)) to collectively represent, as much as possible, the full clinical description of the intervention.

Sample Map:

<table>
<thead>
<tr>
<th>Source Code</th>
<th>Source Code Title</th>
<th>Target Code and extension codes</th>
<th>Target Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDB89</td>
<td>Primary total prosthetic replacement of joint of finger</td>
<td>MGJ.ML.AA</td>
<td>Arthroplasty of finger or hand joint</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MGJ</td>
<td>Joint of hand or fingers</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ML</td>
<td>Reconstruction</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AA</td>
<td>Open approach</td>
<td></td>
</tr>
<tr>
<td>XA9291</td>
<td>Interphalangeal joint of the hand</td>
<td>XDE8</td>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>

Source: KVÅ Version 2018
Target: ICHI Beta Version 2018
<table>
<thead>
<tr>
<th>Cardinality of Map</th>
<th>Relationship of Source Code to Target Code (Equivalence)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:1</td>
<td>Equivalent</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source Code</th>
<th>Source Code Title</th>
<th>Target Code and extension codes</th>
<th>Target Code Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>QK003</td>
<td>Support or training for carrying out daily routine</td>
<td>SDG.PH.ZZ</td>
<td>Training in carrying out daily routine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SDG.PM.ZZ</td>
<td>Education about carrying out daily routine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SDG.RB.ZZ</td>
<td>Practical support with carrying out daily routine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SDG.RC.ZZ</td>
<td>Emotional support for carrying out daily routine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SDG.PH.ZZ+SDG.PM.ZZ+SDG.RC.ZZ</td>
<td>Training in, education about, practical support with,</td>
</tr>
</tbody>
</table>

14 The description of the intervention is: Information, teaching and support as well as training in carrying out simple or complex and coordinated actions to plan, manage and complete the requirements of day-to-day procedures or duties.
<table>
<thead>
<tr>
<th>Cardinality of Map</th>
<th>Relationship of Source Code to Target Code (Equivalence)</th>
<th>RB.ZZ+SDG.RC. ZZ</th>
<th>emotional support for carrying out daily routine</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:4</td>
<td>Broader</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Additional examples

Relationship of source to target maps:

The degree of equivalence was valued concerning using one or several ICHI stem code(s) (with or without extension codes). The degree of equivalence of equal was only used if the cardinality was 1:1. The broader and narrow degrees of equivalence were used for all cardinalities.

Orthopaedic surgical interventions (50 interventions)

Equivalent 16%, Narrower 48%, Broader 34%; None 2%

Cardinality of maps - 1:1 (68%), 1:2 (20%), 1:3 (10%)

The maps demonstrated the need to add extension codes to ICHI to meet the need for additional information, especially for the orthopaedic surgical interventions.

Functioning interventions (50 interventions)

Equivalent 10%; Narrower 2%; Broader 88%;

Cardinality of maps - 1:1 (20%), 1:2 (22%), 1:3 (22%), 1:4 (22%), 1:5 (12%), 1:6 (2%). The diagnostic interventions in the source classification have higher cardinality than the therapeutic interventions for the functioning interventions (Figure 1).
Figure 1: Number of functioning source interventions for each cardinality

Mapping toolkit:
- KVÅ version 2018
- ICHI beta version 2018
- Draft ICHI guidelines
- Mapping process ICHI

Mapping methodology:
1. Select subsets of KVÅ codes (in Swedish) linked to NCSP+ (with titles in English)
2. Map to the best semantic match to each axis of ICHI and ICHI stem code using the draft ICHI, guidelines and search function.
3. If no semantic match is available, leave the target code blank and add “not possible” to map.
4. Select relevant extension code(s) if needed to specify the source intervention
5. Multiple ICHI stem codes (and relevant extension codes) are added according to rules identified in mapping principles.
6. Determine the degree of equivalence for mapping each axis, stem codes, extension codes, and final ICHI stem code with or without extension codes.
7. Show the cardinality of the relation between the source interventions mapped to target intervention(s).

Quality assurance/validation:
- Two mappers performed the mapping
- If both maps created by the two mappers match exactly, the map only was checked a final time together by the two mappers.
● If there is a discrepancy of any type, the two mappers discussed the map to reach a consensus.

Competencies and skill set for mappers:
Both mappers have extensive knowledge of classifications and clinical practice.

Conclusions
Our experiences of this pilot study are that mapping to each axis of intervention and mapping sections by sections in the source classification improve the quality of maps. Using cardinality (described in numbers) and the degree of equivalence are important factors to understand the relationship between the source and target intervention(s). When mapping interventions from the source classification to ICHI, the rules and guidelines for mapping are important for achieving consistency in the mapping, especially regarding cardinality and equivalence.

Appendix 9: International Classification for Nursing Practice to the International Classification of Health Interventions

Authors: Lorraine Block, Gillian Strudwick, Leanne M. Currie, Nicholas R. Hardiker

Source terminology:

International Classification of Nursing Practice (ICNP) (Community Nursing Catalogue)

The International Classification of Nursing Practice is developed and maintained by the International Council of Nurses. It is a multi-hierarchical terminology, with both interface and reference properties. It is a Related Classification within WHO-FIC and conforms to ISO 18104:2014 Health informatics - Categorical structures for the representation of nursing diagnoses and nursing actions in terminological systems. ICNP utilises the Web Ontology Language (OWL) to support automated description logic reasoning. The 2017 release of ICNP comprises 4,326 concepts, representing nursing diagnoses, interventions, and outcomes. ICNP is released in various formats, including interface-oriented subsets such as the ICNP Community Nursing Catalogue.

Target Terminology:

International Classification of Health Interventions (ICHI) Beta 2018

Purpose of this mapping (use cases supported):

There is one use case for this set of maps:

Evaluation of content representation (community nursing interventions) between an agreed subset of interface terms (nursing intervention statements within the ICNP Community Nursing Catalogue) and an administrative classification (ICHI).

This use case will cover the mapping work between the ICNP Community Nursing Catalogue (intervention statements) to the draft 2018 Beta release of ICHI.

Examples of usage:

The ICNP Community Nursing Catalogue contains a subset of terms related to delivering health services in a community context. The map could support the:
➢ Representation of community nursing intervention concepts at a meaningful level of aggregation in ICHI
➢ Collection of community nursing intervention concepts for administrative purposes.

What the maps are:

This is a unidirectional map, where source interface terms (drawn from ICNP) are mapped to a target administrative classification (ICH1). The specific coding methods used to describe the mapping relationships are noted below.

Inputs:
The ICNP input is an agreed subset of community nursing intervention terms (n=187). Each underlying ICNP concept includes a defined action and target property. The ICH1 input is the complete classification. Each ICH1 concept includes a defined target, action and means attribute.

Outputs:
An ICNP concept will be mapped to a fully expressed ICH1 concept (where possible) and include an expressed mapping relationship type.

Relationship types between inputs and outputs:
The semantic match is designated as:

- Broader: ICH1 is broader than ICNP
- Narrower: ICH1 is narrower than ICNP
- Exact match: ICH1 is equivalent to ICNP
- No match: ICH1 has no match to ICNP term.

The map's cardinality is designated as one-to-one, where the source concept was matched to only one target concept.

Mapping attributes:

<table>
<thead>
<tr>
<th>Field</th>
<th>Data Type</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ICNP Intervention Name | String | Composite of numeric ICNP code and ICNP description that identifies the source concept
--- | --- | ---
ICHI Code | String | Identifies the ICHI code that comprises the target concept
ICHI Descriptor | String | ICHI descriptor that identifies the target concept
Matching | String | Identifies the level of equivalence of the map: equivalent, broader, narrower or no equivalence

What the map is not:
This map does not provide alternative and multiple target codes. It only presents the best semantic matches, as determined by the mapping method. For example, cardinality cases of one-to-many and many-to-one were not within scope.

Types of map:
A simple map (cardinality of one-to-one) method was used to link a single ICNP concept to a single fully expressed ICHI concept (where possible).

Sample Map:

<table>
<thead>
<tr>
<th>Source: ICNP</th>
<th>Target: ICHI Beta Version 2017 (or later)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source Code</strong></td>
<td><strong>Source Code Description</strong></td>
</tr>
<tr>
<td>10032483</td>
<td>Promoting Oral Hygiene</td>
</tr>
</tbody>
</table>

Cardinality of Map | Relationship of Source Code to Target Code (Equivalence) | Human-readable Map Advice:
One-to-one | Exact | Mapping decision support facilitated by reviewing targeted ICHI concept definition and hierarchal relationship.

Discussion

This mapping activity demonstrated ICHI had a good representation of community nursing intervention concepts, covering most source terms as equivalent or broader than (which implies the utility of ICHI as an aggregating administrative classification). Further development of ICHI may consider a) adding missing concepts and b) reviewing the semantic degree of separation between ICNP concepts and those broader than/narrower than ICHI matches.

Relationship of source to target maps:

Equivalent 22%; Narrower 5%; Broader 54%; None 19%

Mapping toolkit:

- ICNP release of 2017 (or later as available)
- ICHI Beta 2018 version (or later as available)
- Draft ICHI guidelines
- Spreadsheet to document results

Mapping methodology (manual):

Each coder independently adhered to the following process:

1. Select an ICNP code
2. Enter the concept in the ICHI browser search facility
3. Locate an equivalent semantic match in ICHI. If found, confirm by reviewing the target, action, means axes and using the draft ICHI, guidelines and search function. Record results in mapping spreadsheet.
4. If not found, search for an equivalent semantic match using concept synonyms. If found, confirm by reviewing the target, action, means axes and using the draft ICHI, guidelines and search function. Record results in mapping spreadsheet.
5. If not found, manually search in the ICHI hierarchies for the equivalent ICNP concept. If found, confirm by reviewing the target, action, means axes and using the draft ICHI, guidelines and search function. Record results in mapping spreadsheet.

6. If not found, use the above method (steps 2-5) to find the closest lexical or synonymous match. If found, confirm by reviewing the target, action, means axes and using the draft ICHI, guidelines and search function. Decide if the match is broader than or narrower than. Record results in mapping spreadsheet.

7. If no semantic match or closest match is available, record as a no match in the mapping spreadsheet.

8. Add any human-readable mapping advice for end-users that will be applicable for this map (e.g. synonyms used).

Quality assurance/validation:

- Simple (one-to-one) maps with equivalent, broader than, narrower than, or no match relationships were completed manually by two experienced registered nurses. These coders were additionally experienced in mapping methods (intermediate) and nursing informatics research. Coding occurred in a blinded, parallel fashion.
- Once mapping results were completed, the two coders compared the lists to determine a rate of agreement. Any mismatches between the two coders were reviewed and discussed until consensus was reached (on one agreed target code).
- This process was completed for each batch of concepts (n=10, n=30, n=147). After each phase, results were reviewed by two expert informatics researchers to ensure consistency in process and method.

Release cycles and maintenance:

Mapping cycles to be determined

Drivers for maintenance of existing maps:

- ICHI update cycle
- ICNP update cycle
- Updates to mapping principles
• Additional use cases

Competencies and skill set for map specialists:

• Experienced registered nurses with nursing informatics graduate degrees and experience with mapping methods
• Knowledge of both the source and target classifications and rules for use

Skill level within the team:

• Two coders: intermediate skill level (at least one mapping project completed)
• Two expert informatics researchers: advanced skill level (experience of health terminology development and multiple mapping projects completed)