

International Journal of Population Data Science

Journal Website: www.ijpds.org



Mapping different versions of the international classification of diseases: a scoping review

Nasiba M. Ahmed¹, Mina Nouredanesh¹, Lisa M. Lix¹, and Amani F. Hamad^{1*}

Submission History

Submitted:	23/08/2025
Accepted:	21/01/2026
Published:	12/03/2026

¹ College of Community and Global Health, University of Manitoba, Winnipeg, Manitoba, Canada R3E 0W3

Abstract

Introduction

The International Classification of Diseases (ICD) is used to report medical diagnoses and procedures in most electronic health databases. Periodic revisions and region/country-specific adaptations reflect advances in medical knowledge and local needs. Therefore, health databases often include multiple ICD versions. Comprehensive and accurate mapping between different ICD versions and adaptations is necessary to facilitate research over time and across regions and populations.

Objective

This scoping review describes published literature about methods to map diagnosis codes between different versions or adaptations of the ICD system.

Methods

A systematic search across MEDLINE, EMBASE, Scopus and Web of Science Core Collection from inception until June 21, 2024, was performed. Primary research and review articles describing the mapping methods were included. Information was extracted on article characteristics, data characteristics (e.g., versions of ICD codes mapped), mapping methods, and measures to assess mapping quality. Study data were descriptively analysed using frequencies and percentages.

Results

Among 1359 articles identified from the search, 25 were included in the review; 13 (52%) articles were published between 2020 and 2023. Two-thirds of the articles used US electronic health databases. Mappings were most frequently (48%) created between ICD-9-CM (i.e., 9th revision, clinical modification) and ICD-10-CM (i.e., 10th revision, clinical modification). Mapping methods were mostly (88%) manual, relying primarily on expert review and existing resources to generate mappings and assess their quality. Only 36% of the articles implemented at least two or more strategies to mitigate loss of information due to mapping.

Conclusion

We identified several methods to map different ICD version, all methods relied on expert review to assess accuracy of the mappings. The application of automated approaches through the utilization of AI tools could represent an opportunity for future research.

Keywords

diagnosis codes; code translation; review study

*Corresponding Author:

Email Address: amani.hamad@umanitoba.ca (Amani F. Hamad)



Introduction

The International Classification of Diseases (ICD) is an international standard for recording diagnoses and procedures [1] developed and maintained by the World Health Organization (WHO). ICD revisions are essential for capturing advances in medical science and changes in understanding or interpretation of diseases and conditions. Therefore, multiple versions of the ICD system have been developed. ICD versions 6 (introduced in 1948) through 9 (introduced in 1977) follow the same general framework, but a significant change in structure occurred with ICD-10, which was introduced in 1993. Many countries have adapted the original ICD versions to their specific needs. For example, a clinical modification of the 9th revision (i.e., ICD-9) was created by the US (i.e., ICD-9-CM). Similarly, ICD-10 was modified to ICD-10-CM. Australia, Canada, Germany and other countries have adapted their own versions of ICD-10. For example, ICD-10-CA (i.e., Canadian adaptation) was developed by the Canadian Institute of Health Information (CIHI) to meet Canadian health data needs. This resulted in a large and complex system of codes in use across most jurisdictions and over time.

ICD codes are commonly used to document diagnoses in electronic health databases, such as in administrative health records. Electronic health databases are widely used to conduct policy-relevant health research and surveillance, and are highly advantageous for longitudinal studies about disease and healthcare use trends over time and across regions and populations. Changes and variations in ICD versions found in electronic health databases impact disease definitions and coding practices, thereby affecting how health outcomes are measured and compared. Harmonisation of ICD codes is an essential step to facilitate longitudinal and multi-jurisdictional research.

Mapping, the process of establishing correspondence between different ICD versions or adaptations, is essential to support research using electronic health databases. The outcomes of the mapping process are crosswalks, which document codes that are equivalent or nearly equivalent in different ICD versions. For example, CIHI plays a central role in maintaining national data standards and coding guidance for Canadian administrative health data [2]. ICD crosswalk files developed by CIHI define the correspondence of ICD-10-CA and ICD-9-CM codes [3]. The US Centers for Medicare and Medicaid Services (CMS) and the Centers for Disease Control and Prevention (CDC) introduced general equivalence mappings (GEMs), a publicly available resource to map ICD-9-CM to ICD-10-CM codes [4, 5]. Due to differences in structure and granularity, few direct matches (or one-to-one mappings) exist between the ICD coding systems; a single code from older versions of ICD is often represented by multiple specific codes in the newer versions, resulting in one-to-many mappings. When mapping is performed in a backward direction from a newer version to an older version, often no direct suitable match can be obtained, leading to one-to-zero maps. Both one-to-many and one-to-zero maps present ambiguous mappings; they are more frequent when mapping is done in both directions from older to newer ICD coding versions. Additionally, when retrospective mapping is performed from one ICD coding system to another, a known and inherent challenge is the loss of information. It occurs when the target

ICD system cannot preserve the meaning of the source code, either because no equivalent code exists (i.e., one-to-zero) or because the source code is too broad to map uniquely to one of several more specific target codes (i.e., one-to-many). Thus, while assessing the quality of mappings, accounting for information loss is crucial.

Methodological variations in mapping approaches pose challenges to maintaining semantic equivalence, eventually leading to loss of information, ambiguity and inconsistent mappings depending on the direction of mapping. As countries and health systems begin transitioning to ICD-11, the need for collation and comparison of the existing methods is essential to inform mapping of future versions and to explore the use of AI in mappings. While many studies have developed and evaluated individual mapping methods between ICD versions [6–8], there has been little, if any, assessment of these methods for mapping and the quality of the mappings. Accordingly, we undertook a scoping review to describe published literature about methods to map diagnosis codes between different versions or adaptations of the ICD system. A preliminary search of MEDLINE, the Cochrane Database of Systematic Reviews, JBI (Joanna Briggs Institute) Evidence Synthesis and Open Science Framework was conducted, and no current or underway systematic reviews or scoping reviews on the topic were identified.

Methods

The scoping review was conducted in accordance with the JBI methodology for scoping reviews [9]. A priori protocol was registered in Open Science Framework (<https://osf.io/mktz4/>), which included information regarding the objectives, inclusion criteria and methods for the scoping review.

Search strategy

A comprehensive search of the literature was conducted in MEDLINE, EMBASE, Scopus and Web of Science Core Collection from database inception until June 21, 2024. The search strategy, formulated by the research team and guided by a reference librarian, included concepts related to mapping or crosswalks and ICD. The selection of search terms and keywords was informed by an initial search of MEDLINE to identify relevant articles. Keywords from the titles and abstracts of relevant articles, along with the index terms used to describe them, were used to develop the search strategy. The detailed search strategy implemented for each database is provided in Supplementary Appendix 1a.

Inclusion and exclusion criteria

Peer-reviewed journal articles, including both original research and review articles (e.g., systematic, scoping, or narrative reviews), were included. Conference proceedings, reports and book chapters were excluded. No time or geographical restrictions were applied.

Articles were included if they identified the methods used to map between different versions of ICD codes (i.e., manual and automatic methods). Eligible articles were required to

report on (i) how the mapping was performed, including the mapping logic, algorithm or rules to conduct the mapping, resources used in the mapping process, and (ii) evaluation procedures to assess mapping quality. Manual mapping refers to the process of aligning diagnostic or procedural codes across different versions of the ICD coding system through expert review. It relies on the expertise of trained professionals to assess semantic equivalence, resolve ambiguities, and account for differences in coding structure and specificity. Manual mappings are version-specific, not available for all versions and jurisdiction-specific adaptations. Automated mapping leverages artificial intelligence (AI) techniques such as natural language processing (NLP) or algorithmic reasoning engines; it infers semantic relationships and generates crosswalks between ICD versions without human intervention [10]. Automated mapping methods can handle large volumes of data quickly [11]. Articles that used a hybrid method, which applies both manual and automated methods to either (i) identify candidate mappings, (ii) generate an ICD crosswalk from candidate mappings, or (iii) resolve ambiguous (one-to-many and one-to-zero) mappings, were also included in this review. Finally, articles that conducted mapping between local ICD adaptations (e.g., ICD-10-GM, ICD-10-AM) were included.

Articles were excluded if they met one or more of the following criteria: (1) mapping was between unstructured texts and clinical or medical terms into ICD versions; (2) mapping was between one or more ICD versions to different classification systems (e.g., SNOMED); (3) the article was not in English; (4) the article did not report the mapping method in sufficient methodological detail to allow for data extraction.

Sources of evidence screening and selection

Following the search, all identified citations were exported and uploaded into Covidence, a web application to manage and organise articles through the process of deduplication, title, abstract and full-text screening [12]. A pilot test for titles and abstracts screening was conducted by two reviewers on 50 articles to ensure that the inclusion and exclusion criteria were applied consistently [13]. Following the pilot test, titles and abstracts were screened by two reviewers independently using the inclusion and exclusion criteria. Full-text screening of all selected articles was conducted independently by two reviewers to determine eligibility for inclusion in the review. Inter-reviewer agreement was assessed using the percentage of agreement; the 95% confidence interval (CI) was also estimated.

Reasons for excluding articles at full-text screening were recorded. Disagreements between the reviewers at each stage of the selection process were resolved through discussion until agreement was reached, or by involving an additional reviewer. The results of the search and the article inclusion process are reported following the reporting guidelines by PRISMA-ScR (Supplementary Appendix 2) and presented in a PRISMA flow diagram (Figure 1) [13].

Data extraction and analysis

A data extraction form was developed, and two reviewers independently performed data extraction piloting on a randomly selected 20% of the included articles to evaluate

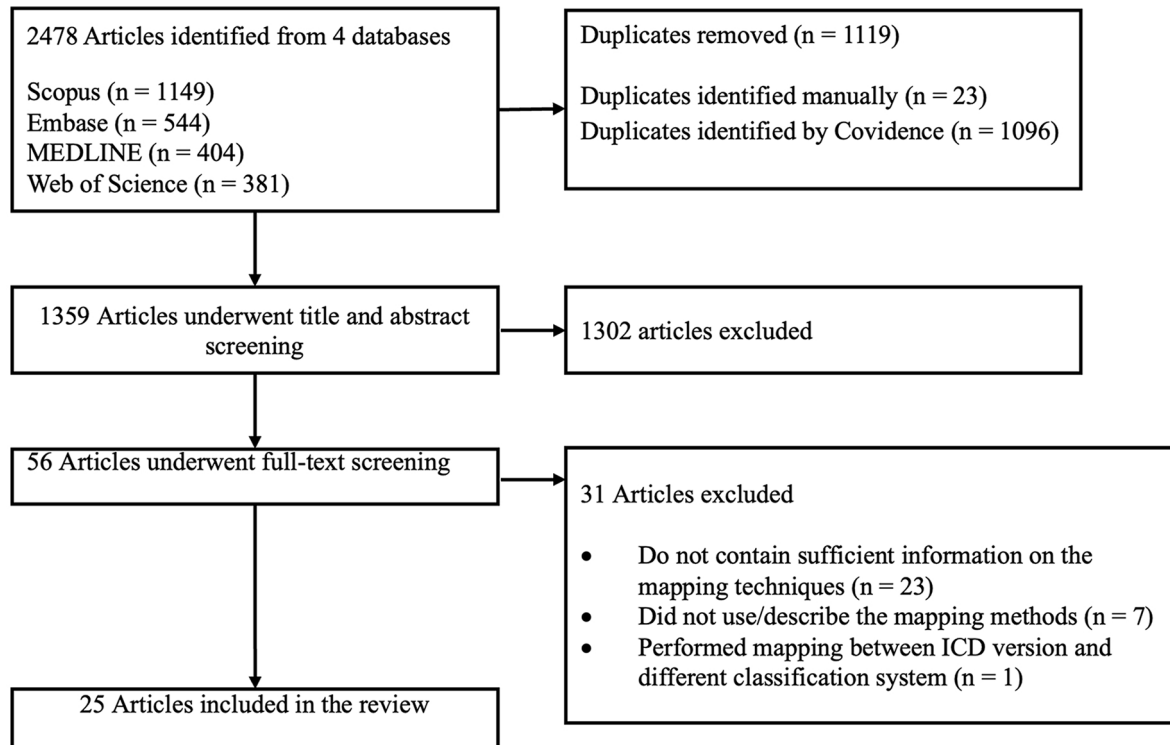
the form's reliability and clarity by calculating inter-reviewer agreement [13]. Discrepancies of the extracted data were resolved by discussion. Percentage agreement and its 95% CI were calculated. Variables with open-ended responses were excluded from this calculation.

The extracted data included article characteristics, such as year of publication, journal discipline (determined based on broad subject terms from the United States National Library of Medicine catalogue), and article type (i.e., primary research, review). Data characteristics that were extracted included geographic location of the data sources, ICD versions of the source and target code, and the topic area of the mapping, which refers to the disease or health condition that was the focus of the mapping. The ICD version from which the mapping originates is referred to as the source code, while the version to which the mapping is directed is referred to as the target code.

The characteristics of the methods, including the type of mapping methods (i.e. manual, automated, hybrid, direction (forward – older version to newer version, backward – newer version to older version, both), mapping logic, whether or not publicly available mapping resources like GEMs or reimbursement mappings (RMs) were used [4, 5], and handling of ambiguous mappings (one-to-many and one-to-zero), were extracted to assess the range and complexity of the methods. Mapping methods would be considered automated based on the 'method' of mapping, regardless of the evaluation technique, which may or may not involve expert validation. Automated application of manual mapping resources (e.g., GEMs) will still be classified as a manual mapping method, as the mappings are not AI- or algorithm-driven. Mapping logic was categorised as expert knowledge-based, rule-based, probabilistic, lexical similarity-based, ontology-based, a combination of two or more, and others [14]. Rule-based mapping applies predefined, explicit rules to establish links between ICD versions. Lexical similarity-based mapping logic refers to surface-level text matching between the label/title of the ICD codes [15]. Ontology-based mapping logic uses structured concept hierarchies to enable the semantic alignment of concepts across coding systems [14].

Data on measures reported to evaluate the quality of mappings were extracted. This included the different types of evaluation methods such as expert validation, descriptive analysis, quantitative measures (recall or precision); use of dual-coded data, extracted as 'yes', 'no' and 'not reported' format. Dual coding is the practice of coding the same record (or the same clinical information) using two ICD versions (e.g., ICD-9-CM and ICD-10-CM). The strategies reported in the articles to address loss of information were classified as (i) no strategy (indicating no steps were taken), (ii) basic (e.g., expert review, clinical validation), and (iii) comprehensive strategy (combination of two or more). The study purpose reported in the article, which was categorised as surveillance, research, clinical, billing, and others, was also extracted. This extraction helped demonstrate the variation in the application of the range of mapping methods, handling of ambiguous mappings and assessment of the quality of mappings, with respect to the purpose. Descriptive analysis was performed using frequencies and percentages.

Figure 1: Scoping review PRISMA flowchart



Results

Search results

The search identified 2478 articles from the four databases. Following de-duplication, 1359 unique articles were identified. Title and abstract screening yielded 56 articles for full-text screening. A total of 25 articles were included in the review for data extraction. Article selection and reasons for exclusion are summarised in Figure 1. The excluded articles were conference proceedings, editorials and case studies ($n=23$), did not describe the mapping method ($n=7$), and performed mapping between ICD versions and another classification system ($n=1$). All the selected articles were based on primary research except for one review [16]. Reviewer agreement (%) for the title and abstract screening was 98.0% (95% CI: 97.2%, 98.7%), for the full-text screening was 83.9% (95% CI: 76.0%, 91.8%), and for the piloting phase of data extraction, it was 82.2% (95% CI: 74.3%, 90.1%).

Article characteristics

Table 1 summarizes the characteristics of the included articles. Approximately 90% ($n=22$) of the articles were published in or after 2010, with 52% ($n=13$) of them published between 2020 and 2023. The geographic location of data sources in 64% ($n=16$) of the articles was the US. Other geographical locations include Canada, South Korea and Portugal. Overall, a large majority of the articles were published in health informatics/medical informatics journals.

Mapping topics and ICD version

Only 12% ($n=3$) of the articles mapped the entire list of codes from all chapters; this work was done for version ICDA-8 to ICD-9 [17], from Danish ICD-8 to Danish ICD-10 [8] and from ICD-9 to ICD-10 [6]. Most articles mapped a subset of ICD codes, often for specific health conditions, such as pain-related conditions [18]. ICD codes for chronic conditions were considered in three articles. Other conditions that were the focus of code subset mapping studies included oncologic conditions, cardiovascular and cardiac health conditions, haematologic conditions, and diabetes. The condition(s) were not specified in 19% ($n=6$) of the articles. A summary of the topic areas of mapping, the source and target ICD coding versions is presented in Table 2.

ICD-9-CM was the source code in 60% ($n=15$) of the included articles. The second most common version was ICD-9, followed by ICDA-8. The remaining source code versions included ICD-8, ICD-10 French version, ICD-10-CM, ICD-11, and ICD for Oncology, 3rd Edition (ICD-O3).

In some articles, the source codes were mapped to more than one version of the target codes. The most frequent (45%, $n=13$) version of the target code was ICD-10-CM. The version of target codes with the second highest occurrence (20%, $n=5$) was ICD-10, followed by ICD-10-PCS, which is used to report procedures in the US. Additional mapped target code versions included ICD-9, ICDA-8, ICD-10-CA, ICD-10 Danish version, ICD-11 and Korean Classification of Diseases, 7th revision or KCD-7. The number of digits considered for mapping in the source codes and target codes was not identified in 13 and 14 articles, respectively. At least four digits

Table 1: Characteristics of the articles included in the scoping review (Count = 25)

Publication decade	Count	Percentage (%)
1980 – 1989	2	8
1990 – 1999	1	4
2000 – 2009	0	0
2010 – 2019	9	36
2020 – 2023	13	52
Geographic location of data		
Canada	2	8
Europe	5	20
US	16	64
Other	2	8
Journal discipline		
Clinical medicine	6	24
Epidemiology	4	16
Health informatics/ Medical informatics	11	44
Health services	2	8
Public health/Population health	2	8

Table 2: Summary of mapping topics and ICD versions in the scoping review articles

Characteristics	%	Article Reference
Topic area of mapping		
Entire list	10	[8]
Chronic health conditions	10	[7, 16, 19]
Cardiovascular and cardiac health conditions	10	[19–21]
Diabetes	6	[19, 22]
Oncologic conditions	13	[20, 22–24]
Haematologic conditions	6	[22, 24]
Other conditions	26	[18–20, 25–29]
Not specified	19	[5, 30–33]
Source coding system		
ICDA-8	8	[17, 29]
ICD-9	12	[6, 22, 25, 34]
ICD-9-CM	60	[5, 7, 16, 18–21, 24, 26–28, 32, 35, 36]
Others	20	[8]
Target coding system		
ICD-9	7	[17, 29]
ICD-10	20	[6, 21–23, 25]
ICD-10-CM	45	[5, 16, 18–20, 24, 26–28, 31, 32, 35, 37]
ICD-10-PCS	14	[20, 26, 36, 37]
Other coding system	14	[8]

were mapped for both source and target codes in 44% (n = 11) of the articles.

Mapping methods

The mapping methods identified in this review were predominantly manual methods (88%, n = 22), primarily relying on expert knowledge for mapping between different ICD versions. At least two experts were involved in the manual mapping process. Each mapping method included review and selection of the relevant source and target codes as per the

corresponding objective of the study. Candidate target codes that best matched the group of source codes were identified, followed by the generation of the crosswalk. A substantial majority of the manual mapping methods used publicly available resources that were based on or derived from GEMs to implement manual mappings. GEMs provide one-to-one and one-to-many mappings. On the other hand, reimbursement mappings (RMs), also maintained by CMS, are an applied mapping of the ICD-10 to ICD-9 GEMs [5, 38]. The RMs provide one-to-one mappings for obtaining financial neutrality between ICD-9 and ICD-10 for billing and reimbursement

purposes. Additional mapping resources, primarily based on GEMs, that are reported in the included articles are translation network-based mapping tool [24, 28], assistive encoding software [32], and automated stand-alone mapping tool called MapIT [20, 25]. The network-based mapping captures both direct and indirect relationships between source and target code, including one-to-many and many-to-one mappings [28]. The MapIT and the assistive encoding software offer similar functionality – suggest candidate codes based on GEMs, streamlining the manual mapping method [32].

The mapping logic adopted for manual methods was rule-based, primarily guided by the expert review and often combined with ontology-based [8, 28, 30] and lexical similarity-based [33] logic. The direction of mapping among the manual methods was mostly ($n=11$) in both forward and backward directions. Among all the articles that performed manual mapping, two articles did not report how ‘one-to-many’ or ‘one-to-zero’ mappings were handled [16, 27], and one article excluded these ambiguous mappings by design [8]. Post-coordination, a method to link a stem code with extension codes in a cluster, was used to handle ‘one-to-many’ mappings in only one article [33]. The stem code refers to the core diagnosis, and the extension codes are appended to specify the site, laterality, severity, etc. This system allowed the use of more than one code to fully represent a more specific or complex clinical concept.

As outlined in Table 3, none of the included articles used automated methods, but three articles used hybrid mapping methods based on a combination of manual and automated methods [6, 23, 31]. While the specific methods varied, all the hybrid mapping methods conducted the mappings in multiple stages. The manual component involves the use of existing mapping resources (developed manually) to

generate candidate mappings [23, 31], and expert review to resolve ambiguous (one-to-many and one-to-one) mappings [6, 31]. The existing mapping resources included mapping tables from SNOMED CT, the U.S. National Cancer Institute Metathesaurus (NCI Mt) and the Unified Medical Language System (UMLS) Metathesaurus. The UMLS Metathesaurus is a multi-purpose biomedical database that integrates and standardises terminology from more than 200 biomedical vocabularies, classifications, and coding systems, facilitating semantic interoperability across diverse healthcare information systems [39]. The NCI Mt is a specialised version of the UMLS Metathesaurus, focused on biomedical and cancer-related terminologies [39]. The automated components of hybrid methods applied NLP to obtain candidate mappings [6, 31] and ELK reasoner, an algorithm used to infer logical relationships in ontologies based on description logic, to refine the mappings using SNOMED CT’s logical structure from the candidate mappings [10, 23].

All the hybrid methods utilised a combination of at least two or more mapping logics, including rule-based NLP, lexical-based similarity and ontology-based logic. The mapping was done simultaneously in both forward and backward directions. These methods accommodated one-to-many mappings and one-to-zero mappings. Although no additional steps were taken to refine one-to-zero mappings, all three hybrid methods addressed one-to-many mappings by either allowing the specific codes to fallback to parent code, or by using post-coordination system and expert review.

A summary of the mapping challenges and strategies to handle ambiguity and loss of information is provided in Supplementary Appendix 1c. Among the included articles, the ambiguous mappings like one-to-zero were either excluded or not reported at all in 48% ($n=12$) of the articles, whereas

Table 3: Summary of the methods used for mapping and their quality assessment

Characteristics	%	Article Reference
Mapping Methods		
Manual	88	[5, 7, 8, 16–22, 24–30, 32–36]
Automated	0	-
Hybrid	12	[6, 23, 31]
Use of publicly available mapping resources		
Yes (GEMs, RMs)	68	[5, 16–22, 24–28, 32, 34–36]
No	32	[6–8, 23, 29–31, 33]
Assessment of mapping quality		
Expert validation	28	[8, 17, 19, 20, 23, 24, 36]
Descriptive analysis	4	[5]
Trend analysis	20	[18, 26–28, 35]
Expert validation & descriptive analysis	24	[6, 7, 21, 31, 33, 34]
Expert validation & trend analysis	16	[16, 22, 25, 29]
Others (quantitative measures: comparability factors and ratios, recall and precision)	8	[30, 32]
Strategy to address loss of information		
No strategy	48	[5, 7, 16–19, 21, 22, 27, 28, 32, 35]
Basic strategy (expert review, clinical validation)	16	[24, 26, 29, 36]
Comprehensive strategy (combination of two or more)	36	[6, 8, 20, 23, 25, 30, 31, 33, 34]

one-to-many mappings were included in most of the articles (88%, $n = 22$). Additional steps to mitigate the loss of information due to the ambiguous mappings (i.e., one-to-zero and one-to-many) were taken in only 16% ($n = 4$) of the articles.

Assessment of the quality of mapping

All the methods identified from the articles included in this review employed at least one evaluative measure, for instance, expert validation, descriptive analysis, and trend analysis, to assess the quality of mappings. Around 28% of the articles ($n = 7$) relied solely on expert validation, and more often in combination with descriptive analysis (24%, $n = 6$) and trend analysis (16%, $n = 4$). The descriptive measures included percentage calculation of one-to-one, one-to-many and one-to-zero mappings, quality of fit and concordance rate (degree of agreement between the source and target codes). Trend analysis was independently used in 20% ($n = 5$) of the articles to examine the effects of the transition of the ICD coding systems. Healthcare metrics used in the trend analysis include proportions, incidence and prevalence rates of specific conditions, total number of admissions due to specific conditions, number of hospital episodes, and proportion of diagnosis codes. The observed trends before and after transition were found to be consistent in the majority of the studies reporting the findings on trend analysis [18, 25, 27, 29, 35], while inconsistency was observed in some of them [16, 22, 26, 28].

Only three articles, all employing manual methods, used dual-coded data for assessing the validity of the mappings [5, 30, 32]. The use of dual-coded data facilitated the calculation of quantitative measures like recall [30], precision [30] and comparability factors [32]. Recall was defined as the percentage of correct target codes in the reference standard that were identified by the mapping method. Precision was defined as the percentage of correct target codes among all the target codes identified by the method. A precision of 80.2% and a recall of 70.4% was reported in the only article that computed quantitative measures [30]. Comparability factor, developed by the National Center for Health Statistics, quantifies the effect of changes in coding rules and classification in the ICD revision [40]. Hence, the measure offers an understanding of the continuity of longitudinal data between different ICD coding systems. A comparability factor of 100 indicates minimal discontinuity, less than 100 indicates fewer cases coded in the newer version than in the older version and vice versa. The article reporting on the comparability factor indicated that the mappings yielded a range of values (from 16.2 to 118.0), varying by the specific conditions studied. The purpose of both articles was research for clinical decision support [30] and surveillance of longitudinal trends in public health [32]. Despite using dual-coded data, one of these three articles conducted descriptive analysis (percentage of one-to-one and one-to-many mappings) to assess the validity of the findings; it conducted mapping for billing purposes [5].

The loss of information due to mapping between different ICD coding systems was acknowledged and reported in all the articles, although a significant proportion (48%, $n = 12$) adopted no additional strategy to mitigate the loss of

information, irrespective of the purpose of the study. A basic strategy, like expert review, was employed in some articles (16%, $n = 4$). A manual mapping method was used in all of these articles discussed above. A combination of at least two strategies was implemented in all the articles that employed the hybrid method, whereas only a subset of the articles applying manual methods adopted such a comprehensive approach to address loss of information. The primary purposes of mapping in all these articles (36%, $n = 9$) that attempted to minimise the loss of information with multiple strategies were clinical research for decision support and public health surveillance. The additional strategies include categorisation of mapping complexity, allowing specific codes to fallback to parent codes by truncation, the use of pragmatic mappings and residual categories like 'other categories' or 'unspecified'. Under the assumption of minor impact of misclassification and when avoiding one-to-zero mapping altogether may lead to greater loss of information, pragmatic mappings assign the source code to the closest or most reasonable target code [6]. Relatively consistent mappings of ICD codes were found in the majority of the articles [6, 23, 25, 30, 31, 34] that utilised comprehensive strategies to minimise loss of information, in comparison to articles that did not address it [16, 22, 28, 32]. The detailed characteristics of the mapping methods and validation methods reported in the articles are available in Supplementary Appendix 1b.

Discussion

Our scoping review aimed to synthesise the current landscape of mapping methods between different versions or adaptations of the ICD coding system. We identified 25 eligible articles across four major electronic databases that conducted mapping between different versions or adaptations of ICD codes, adopting either manual or hybrid methods. The majority of the articles were based in the US. We identified a few articles that used data sources from European countries and Canada, which collect and maintain comprehensive administrative databases. ICD code mapping workflows are often shaped by local coding guidelines, reimbursement structures, and health-system governance. Most articles (52%) were published in or after 2020 and predominantly mapped between ICD-9-CM and ICD-10-CM coding systems. A significant majority of the articles conducted mapping using a manual method and relied on pre-existing mapping resources. Both manual and hybrid methods mostly included one-to-many mappings, whereas the one-to-zero mappings were either excluded or included without any additional refinement. Dual-coded data was successfully utilised in only two articles, which facilitated the computation of quantitative measures to assess the quality of mappings. Loss of information was addressed by adopting at least one strategy in 52% (Table 3) of the articles, while a comprehensive strategy, a combination of two or more strategies, was implemented in the articles with clinical research for decision support and public health surveillance purposes.

Our review identified articles that performed mapping between the structured coded versions of ICD, generating code-to-code crosswalks between the specific versions or adaptations. The review did not include articles that

predominantly leverage AI-based NLP and machine learning approaches to conduct mapping between unstructured texts (originating from clinical notes or discharge summaries) and the ICD coding system [41, 42]. Thus, no article in this review was identified that employed an automated method, which could generate entirely AI-driven or algorithm-driven mappings without any human involvement. One of the potential barriers that limit the application of AI-based, automated methods to map ICD codes is the variation in the structure, hierarchy, and/or classification principles of coding systems between different ICD versions. In addition, the scarcity of dual-coded data is another fundamental barrier to the development and validation of automated ICD code mapping methods. Dual-coded data can facilitate training of automated methods and evaluation of mapping quality (e.g., recall, precision, and comparability ratios). This review shows that dual-coded datasets are not frequently used in ICD code mapping; this can be attributed to the additional coding workload and costs required to maintain dual coding during transition periods. Moreover, administrative health data systems often rely on nationally endorsed crosswalks (e.g., GEMs, the CIHI crosswalk table) for reporting and reimbursement. Hence, research organisations often continue to use these tools rather than experiment with new automated mapping methods. This may limit opportunities for the development and evaluation of automated methods in real-world settings and their reporting in the peer-reviewed literature. As identified in this review, a substantial majority of the articles applying either manual or hybrid methods used pre-existing mapping resources, such as GEMs. The use of GEMs alone may be insufficient for accurately determining ICD-10-CM crosswalks from ICD-9-CM [19, 25], as GEMs were primarily designed for billing rather than clinical research. Literature shows that the use of GEMs for ICD code mapping may often lead to misrepresentation of certain populations and/or procedures [5]. For instance, in one of the articles included in this review, codes from dual-coded records were compared with codes generated using the GEMs and RMs [5]. Although most mappings were reproducible, many one-to-one mappings were classified as approximate, with mismatches reflecting differences in coder interpretation and code hierarchy [5].

The latter versions of ICD coding systems consist of more specific and detailed codes than previous versions. Therefore, ambiguous mappings like one-to-one or one-to-many were reported in almost all the articles. A pattern on how one-to-many or one-to-one mappings were handled could not be conclusively ascertained. Additionally, the differences in structure, quality of coding in clinical practice, and lack of a consistent residual category system (classes named “other” and “unspecified”) across different versions of ICD coding systems pose significant challenges in the fully automated conversion between ICD versions [6]. The majority of the articles did not report on how the loss of information was addressed when a source ICD code had no direct equivalent in the target ICD version (i.e., one-to-zero mapping). To improve transparency and comparability, we recommend that future mapping studies report on (i) the presence and proportion of one-to-many and one-to-zero mappings, (ii) the criteria used to handle such ambiguous mappings, and (iii) description of any steps/strategies taken to minimise the loss of information. This can facilitate better understanding of the degree of

information loss introduced during mapping, more importantly for the researchers using longitudinal health data. Niyirora (2021) proposed a quantitative measure that can provide an understanding of the complexity in transitioning between different ICD coding systems [37]. It was demonstrated that the measure can help identify clinical concepts that are prone to higher documentation and coding challenges during mapping transitions.

As there is no universally accepted “correct” match in ICD mappings in the absence of dual-coded data, the assessment of the quality of mappings mostly included trend analysis and percentage calculation of structural indicators, such as one-to-one and one-to-many mappings. Ideally, the transition from one coding system to another should not distort true disease patterns or trends. Hence, accurate mappings should yield the same number of cases for any condition (e.g., heart attacks, diabetes) before and after the transition. Notable changes in the pattern/trend after the transition, without any other intervention, can indicate mapping quality issues. On the other hand, a higher percentage of one-to-one mapping is often presumed as an indicator of structurally well-aligned mappings, which may not always be true. The articles that did not take steps to mitigate loss of information while mapping, but did include one-to-many mappings [5, 18, 21, 35], obtained relatively consistent mappings in comparison to articles that excluded one-to-many mappings. Dual-coded data can provide a real-world reference while assessing mappings between different versions of ICD. Although only two articles in this review demonstrated the strategic use of such data, it can also be used as a training and test dataset while developing an automated method for mapping, as the dual-coded records will exemplify real-world mappings between the source and target ICD codes [43]. It would necessitate accounting for the variation in coding practices, coder proficiency levels and representativeness of the dataset. While mapping between different versions or adaptations of ICD coding systems, loss of information is inherent. Regardless of the method used, manual or hybrid, consideration of a comprehensive strategy to minimise the loss is crucial to generate mappings with better accuracy.

Although manual mapping of ICD codes is still considered the gold standard, a combination of ontology-based semantic integration, which is the process of aligning and merging different terminologies or classification systems based on their semantics, with existing automated approaches can be helpful to limit the manual component [23, 44]. This approach may rely on structured knowledge bases (e.g., SNOMED CT) as intermediaries, linking related concepts through hierarchies, synonyms, and predefined mappings. Thus, this approach can preserve the clinical meaning and relationships while mapping. Additionally, exploration of AI-based approaches like NLP, especially biomedical-specific language models and the utilisation of the descriptions/title or labels of ICD codes, is a promising avenue to achieve automated ICD conversions [45]. This may result in more semantically accurate mappings between ICD versions with reduced manual effort, in a time-sensitive manner, while addressing the ambiguities arising from one-to-many or one-to-zero mappings.

One of the strengths of our review is the systematic, robust approach implemented in the search and screening of the articles for inclusion, without imposing any restrictions

on the timeframe or geographical location. Predefined search terms guided by a reference librarian, and broad eligibility criteria were applied to ensure the selection of relevant articles. Consensus achieved through two rounds of screening (title-abstract and full-text) and a pilot extraction of articles by two reviewers minimised the selection and extraction bias.

Our review has some limitations. Our search may have excluded relevant articles indexed in other databases. This is unlikely to significantly impact the findings, as the selected databases collectively cover the vast majority of peer-reviewed literature, ensuring that the most relevant articles describing mapping methods were included in the review. Due to the challenges associated with translation and time and resource constraints, grey literature was not searched, which may have resulted in the exclusion of relevant conference proceedings, government reports, or organizational newsletters. Moreover, the majority of the articles included in this review were conducted in settings with different levels of national oversight for coding standards. For example, in Canada, CIHI plays a central role in setting national coding standards and overseeing data quality [2], whereas in the US, coding standards and practices are developed and implemented by multiple federal agencies, professional organisations, and individual health systems, resulting in a more decentralised structure. We were unable to examine how these structural differences affect mapping practices, as the available literature lacked sufficient detail to compare mapping practices across national coding environments. In this review, we only considered those articles that mapped between different versions of the ICD coding systems. Hence, articles that described mapping between ICD coding systems and other structured medical terminology systems, such as SNOMED CT, were not within the scope of this review. Future reviews could address this limitation by broadening the inclusion criteria to include articles on related mapping tasks where automated methods are more commonly reported.

Conclusion

Our review illustrated frequent utilisation of manual mapping methods and a few hybrid methods. The use of automated methods, such as machine learning approaches, was limited. The integration of AI-driven techniques such as biomedical NLP models, deep learning, and knowledge graphs holds great potential for automating ICD code mapping. Despite the importance of mapping between versions or adaptations of ICD codes in supporting the clinical research and surveillance of public health over time, there remains a lack of guidance on best practices for handling ambiguous mappings (one-to-many and one-to-zero) and loss of information during mapping, and the utilisation of appropriate quantitative measures to assess the accuracy of the mappings.

Acknowledgements

We would like to thank Carol A. Cooke, MLIS (Librarian, NJM Health Sciences Library, University of Manitoba), for guiding the search strategy development and syntax translation.

Funding

Canadian Institutes of Health Research (CIHR) Project Grant (FRN 487890).

Author contributions

NMA: Conceived the study, prepared the analysis plan, performed the literature search, screened for study inclusion/exclusion, extracted data, and prepared the draft manuscript. MN: Performed the screening for study inclusion/exclusion, piloted the data extraction tool and reviewed the manuscript. LML: Obtained funding, conceived the study, prepared the analysis plan, and reviewed the draft manuscript. AFH: Conceived the study, prepared the analysis plan, performed the literature search, screened articles for resolving conflicts, and reviewed the draft manuscript. All authors approved the final version of the manuscript.

Conflicts of interest

There is no conflict of interest in this project.

References

1. World Health Organization. International statistical classification of diseases and related health problems: 11th revision (ICD-11) 2019. <https://www.who.int/standards/classifications/classification-of-diseases>.
2. Canadian Institute for Health Information. Canadian Coding Standards for Version 2022 ICD-10-CA and CCI. Ottawa, ON: 2022. https://secure.cihi.ca/free_products/canadian-coding-standards-2022-en.pdf
3. Canadian Institute for Health Information. ICD-10-CA/CCI to ICD-9/ICD-9-CM/CCP — Conversion tables 2020. <https://secure.cihi.ca/estore/productSeries.htm?pc=PCC1940>
4. Centers for Medicare & Medicaid Services (CMS), Centers for Disease Control and Prevention (CDC). ICD-10 General Equivalence Mappings: An Introduction 2009. <https://www.cms.gov/Medicare/Coding/ICD10/downloads/ICD10MappingFactSheetIntroduction.pdf>.
5. Turer RW, Zuckowsky TD, Causey HJ, Rosenbloom ST. ICD-10-CM Crosswalks in the primary care setting: assessing reliability of the GEMs and reimbursement mappings. *Journal of the American Medical Informatics Association*. 2015;22:417–25. <https://doi.org/10.1093/jamia/ocu028>
6. Schulz S, Zaiss A, Brunner R, Spinner D, Klar R. Conversion problems concerning automated mapping from ICD-10 to ICD-9. *Methods of Information in Medicine*. 1998;37:254–9. <https://doi.org/10.1055/s-0038-1634529>

7. Hamad AF, Vasyilkiv V, Yan L, Sanusi R, Ayilara O, Delaney JA, et al. Mapping three versions of the international classification of diseases to categories of chronic conditions. *International Journal of Population Data Science*. 2021;6:1406. <https://doi.org/10.23889/ijpds.v6i1.1406>
8. Pedersen MK, Eriksson R, Reguant R, Collin C, Pedersen HK, Sorup FKH, et al. A unidirectional mapping of ICD-8 to ICD-10 codes, for harmonized longitudinal analysis of diseases. *European Journal of Epidemiology*. 2023;38:1043–52. <https://doi.org/10.1007/s10654-023-01027-y>
9. Aromataris E, Lockwood C, Porritt K, Jordan Z. JBI manual for evidence synthesis. Joanna Briggs Institute; 2024. <https://doi.org/10.46658/JBIMES-24-01>
10. Kazakov Y, Krötzsch M, Simančík F. The incredible ELK: From polynomial procedures to efficient reasoning with e ontologies. *Journal of Automated Reasoning*. 2014;53:1–61. <https://doi.org/10.1007/s10817-013-9296-3>
11. Sylvestre E, Bouzillé G, McDuffie M, Chazard E, Avillach P, Cuggia M. A semi-automated approach for multilingual terminology matching: Mapping the French version of the ICD-10 to the ICD-10 CM. *Studies in Health Technology and Informatics*. 2020;270:18–22. <https://doi.org/10.3233/SHTI200114>
12. Mellor L, Covidence team. How to quickly complete full text screening in a systematic review 2023. <https://www.covidence.org/blog/how-to-quickly-complete-full-text-screening-in-a-systematic-review/>.
13. Tricco AC, Lillie E, Zarin W, O'Brien K, Colquhoun H, Kastner M, et al. A scoping review on the conduct and reporting of scoping reviews. *BMC Medical Research Methodology*. 2016;16. <https://doi.org/10.1186/s12874-016-0116-4>
14. Almborg A-H, Baker S, Brear H, Celik C, Chute CG, Mea V Della, et al. WHO-FIC classifications and terminology mapping: principles and best practice 2021. https://cdn.who.int/media/docs/default-source/classification/who-fic-network/whofic_terminology_mapping_guide.pdf
15. Laghari A. An overview of methods and tools for biomedical ontology matching 2023. <https://easychair.org/publications/preprint/gznQ/open>
16. Nam YH, Mendelsohn AB, Panozzo CA, Maro JC, Brown JS. Health outcomes coding trends in the US Food and Drug Administration's Sentinel System during transition to International Classification of Diseases-10 coding system: A brief review. *Pharmacoepidemiology and Drug Safety*. 2021;30:838–42. 10.1002/pds.5216.
17. Colls MH. A method for converting a disease registry's case-load to a new classification of diagnostic codes. *Medical Informatics*. 1980;5:121–30. <https://doi.org/10.3109/14639238009014006>
18. Mayhew M, DeBar LL, Deyo RA, Kerns RD, Goulet JL, Brandt CA, et al. Development and assessment of a crosswalk between ICD-9-CM and ICD-10-CM to identify patients with common pain conditions. *The Journal of Pain*. 2019;20:1429–45. <https://doi.org/10.1016/j.jpain.2019.05.006>
19. Simeone JC, Liu X, Bhagnani T, Reynolds MW, Collins J, Bortnichak EA. Comparison of ICD-9-CM to ICD-10-CM crosswalks derived by physician and clinical coder vs. automated Methods. *Perspectives in Health Information Management*. 2021;18:1e. PMID:PMC8120674.
20. Utter GH, Cox GL, Atolagbe OO, Owens PL, Romano PS. Conversion of the Agency for Healthcare Research and Quality's Quality Indicators from ICD-9-CM to ICD-10-CM/PCS: The Process, Results, and Implications for Users. *Health Services Research*. 2018;53:3704–27. <https://doi.org/10.1111/1475-6773.12981>
21. Columbo JA, Kang R, Trooboff SW, Jahn KS, Martinez CJ, Moore KO, et al. Validating publicly available crosswalks for translating ICD-9 to ICD-10 diagnosis codes for cardiovascular outcomes research. *Circulation Cardiovascular Quality and Outcomes*. 2018;11:e004782. <https://doi.org/10.1161/CIRCOUTCOMES.118.004782>
22. He M, Santiago Ortiz AJ, Marshall J, Mendelsohn AB, Curtis JR, Barr CE, et al. Mapping from the International Classification of Diseases (ICD) 9th to 10th revision for research in biologics and biosimilars using administrative healthcare data. *Pharmacoepidemiology and Drug Safety*. 2020;29:770–7. <https://doi.org/10.1002/pds.4933>
23. Nikiema JN, Jouhet V, Mougin F. Integrating cancer diagnosis terminologies based on logical definitions of SNOMED CT concepts. *Journal of Biomedical Informatics*. 2017;74:46–58. <https://doi.org/10.1016/j.jbi.2017.08.013>
24. Venepalli NK, Qamruzzaman Y, Li JJ, Lussier YA, Boyd AD. Identifying clinically disruptive International Classification of Diseases 10th Revision Clinical Modification conversions to mitigate financial costs using an online tool. *Journal of Oncology Practice*. 2014;10:97–103. <https://doi.org/10.1200/JOP.2013.001156>
25. Dalton MK, Sokas CM, Castillo-Angeles M, Semco RS, Scott JW, Cooper Z, et al. Defining the emergency general surgery patient population in the era of ICD-10: Evaluating an established crosswalk from ICD-9 to ICD-10 diagnosis codes. *Journal of Trauma and Acute Care Surgery*. 2023;95:899–904. <https://doi.org/10.1097/TA.0000000000004050>
26. Tian Y, Ingram M-CE, Raval M V. A pitfall of using general equivalence mappings to estimate national trends of surgical utilization for pediatric patients. *Journal of Pediatric Surgery*. 2020;55:2602–7. <https://doi.org/10.1016/j.jpedsurg.2020.03.011>

27. Ascencao R, Nogueira P, Sampaio F, Henriques A, Costa A. Adverse drug reactions in hospitals: population estimates for Portugal and the ICD-9-CM to ICD-10-CM crosswalk. *BMC Health Services Research*. 2023;23:1222. <https://doi.org/10.1186/s12913-023-10225-z>
28. Noorbakhsh KA, Berger RP, Ramgopal S. Comparison of crosswalk methods for translating ICD-9 to ICD-10 diagnosis codes for child maltreatment. *Child Abuse & Neglect*. 2022;127:105547. <https://doi.org/10.1016/j.chiabu.2022.105547>
29. Colliver J, Dufour M, Bertolucci D, Van Natta P, Malin H. A system to convert ICD diagnostic codes for alcohol research. *Alcohol Health and Research World*. 1984;9:53–5. PMID:6423956.
30. Xu J, Fung KW, Bodenreider O. Sequential Mapping - A novel approach to map from ICD-10-CM to ICD-11. *Studies in Health Technology and Informatics*. 2022;290:96–100. <https://doi.org/10.3233/SHTI220039>
31. Sylvestre E, Bouzille G, McDuffie M, Chazard E, Avillach P, Cuggia M. A semi-automated approach for multilingual terminology matching: mapping the French version of the ICD-10 to the ICD-10 CM. *Studies in Health Technology and Informatics*. 2020;18–22. <https://doi.org/10.3233/SHTI200114>
32. Fenton SH, Benigni MS. Projected impact of the ICD-10-CM/PCS conversion on longitudinal data and the joint commission core measures. *Perspectives in Health Information Management*. 2014;11:1g. PMID:25214824.
33. Lee H. Mapping ICD-11 (The 11th International Classification of Disease) to ICD-10-KM-7th (The Korean Modification 7th of the ICD-10) for flexible transition to ICD-11. *Perspectives in Health Information Management*. 2021;18:1b. PMID:34858114.
34. Hernandez-Ibarburu G, Perez-Rey D, Alonso-Oset E, Alonso-Calvo R, de Schepper K, Meloni L, et al. ICD-10-CM extension with ICD-9 diagnosis codes to support integrated access to clinical legacy data. *International Journal of Medical Informatics*. 2019;129:189–97. <https://doi.org/10.1016/j.ijmedinf.2019.06.010>
35. Panozzo CA, Welch EC, Woodworth TS, Huang T-Y, Her QL, Gagne JJ, et al. Assessing the impact of the new ICD-10-CM coding system on pharmacoepidemiologic studies-An application to the known association between angiotensin-converting enzyme inhibitors and angioedema. *Pharmacoepidemiology and Drug Safety*. 2018;27:829–38. <https://doi.org/10.1002/pds.4550>
36. Hernandez-Ibarburu G, Perez-Rey D, Alonso-Oset E, Alonso-Calvo R, Voets D, Mueller C, et al. ICD-10-PCS extension with ICD-9 procedure codes to support integrated access to clinical legacy data. *International Journal of Medical Informatics*. 2019;122:70–9. <https://doi.org/10.1016/j.ijmedinf.2018.11.002>
37. Niyirora J. Entropic measures of complexity in a new medical coding system. *BMC Medical Informatics and Decision Making*. 2021;21:124. <https://doi.org/10.1186/s12911-021-01485-y>
38. Centers for Medicare and Medicaid Services. ICD-10-CM/PCS to ICD-9-CM reimbursement mappings 2016. <https://www.cms.gov/files/document/reimbursement-mapping-data-2016-1-2.pdf>.
39. Bodenreider O. The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Research*. 2004;32. <https://doi.org/10.1093/nar/gkh061>
40. Anderson RN, Miniño AM, Hoyert DL, Rosenberg HM. Comparability of cause of death between ICD-9 and ICD-10: preliminary estimates. *National Vital Statistics Reports*. 2001;49:1–32. PMID:11381674.
41. Jha J, Almagro M, Tissot H. Designing NLP applications to support ICD coding: an impact analysis and guidelines to enhance baseline performance when processing patient discharge notes. *Journal of Digital Health*. 2023;2:63–81. <https://doi.org/10.55976/jdh.22023119463-81>
42. Xue J, Lu P. ICD code mapping model based on clinical text tree structure. *Artificial Intelligence in Medicine*. 2025;167. <https://doi.org/10.1016/j.artmed.2025.103163>
43. Poltavskiy EA, Fenton SH, Atolagbe O, Sadeghi B, Bang H, Romano PS. Exploring the implications of the new ICD-10-CM classification system for injury surveillance: analysis of dually coded data from two medical centres. *Injury Prevention : Journal of the International Society for Child and Adolescent Injury Prevention*. 2021;27:i19–26. <https://doi.org/10.1136/injuryprev-2019-043519>
44. Wan L, Song J, He V, Roman J, Whah G, Peng S, et al. Development of the International Classification of Diseases Ontology (ICDO) and its application for COVID-19 diagnostic data analysis. *BMC Bioinformatics*. 2021;22. <https://doi.org/10.1186/s12859-021-04402-2>
45. Fang L, Chen Q, Wei C-H, Lu Z, Wang K. Bioformer: an efficient transformer language model for biomedical text mining. *ArXiv [Preprint]*. 2023. PMID:36945685.