ICD-10-PCS extension with ICD-9 procedure codes to support integrated access to clinical legacy data

G. Hernandez-Ibarburu[a,*], D. Perez-Rey[a,*], E. Alonso-Oset[a], R. Alonso-Calvo[a], D. Voets[b], C. Mueller[b], B. Claerhout[b], N.V. Custodix[b]

[a] *Biomedical Informatics Group, Departamento de Inteligencia Artificial, ETSI Informaticos, Universidad Politecnica de Madrid, 28660 Boadilla del Monte, Madrid, Spain* [b] *Sint-Martens-Latem, Belgium*

A B S T R A C T

Since the creation of The International Classification of Diseases (ICD), new versions have been released to keep updated with the current medical knowledge. Migrations of Electronic Health Records (EHR) from ICD-9 to ICD-10-PCS as clinical procedure codification system, has been a significant challenge and involved large resources. In addition, it created new barriers for integrated access to legacy medical procedure data (frequently ICD-9 coded) with current data (frequently ICD-10-PCS coded). This work proposes a solution based on extending ICD-10-PCS with a subgroup of ICD-9-CM concepts to facilitate such integrated access. The General Equivalence Mappings (GEMs) has been used as foundation to set the terminology relations of these inserted concepts in ICD-10-PCS hierarchy, but due to the existence of 1-to-many mappings, advanced rules are required to seamlessly integrate both terminologies. With the generation of rules based on GEMs relationships, 2014 ICD-9 concepts were included within the ICD-10-PCS hierarchy. For the rest of the concepts, a new method is also proposed to increase 1-to-1 mappings. As results, with the suggested approach, the percentage of ICD-9-CM procedure concepts that can be mapped accurately (avoiding mappings to a large number of concepts) rise from 11.56% to 69.01% of ICD-9-Proc, through the extended ICD-10-PCS hierarchy.

1. Introduction

Data storage in Electronic Health Records (EHR) has increased dramatically, but there is a lack of methods to exploit such repositories for clinical research. In this regard, common access to clinical data across institutions with highly heterogeneous information systems is still one of the main challenges. To achieve semantic interoperability, clinical information systems must not only be able to exchange data but also to equally interpret such data. In the process of implementing clinical data interoperability, and due to the different coding systems and data models used, data correspondences, links or "mappings", should be provided among different sources to a common data model and terminology. For the scope of this work, we will focus on the terminology mapping. The purpose of the terminology mapping is to find the synonymous concepts among different terminologies. An example of the terminology mapping is shown in Fig. 1. Clinical data mapping tasks are intended to facilitate that systems are able to correctly use clinical data and also to enable the use of legacy data in modern EHR systems [1].

The International Classification of Diseases (ICD) is a medical terminology developed and revised by the World Health Organization. The predecessor of ICD-10, ICD Ninth Revision Clinical Modification (ICD-9), describes and categorizes diagnoses and procedures associated with hospitals in the United States.

The International Classification of Diseases, Ninth Revision, Clinical Modification volume 3, Procedures (ICD-9-Proc) is the subsection of ICD-9 used for medical procedure codification.ICD-9-Proc has been used since the 80s [2] but the Health Human Services mandated to replace it with ICD-10 Procedure Coding System (ICD-10-PCS) by the end of 2015 [3]. Thus, a lot of effort and work has been spent in the migration of EHR systems from ICD-9-Proc to ICD-10-PCS, trying to reduce as much as possible the costs of this process in terms of money, time and legacy data accuracy [4].

In this paper, we propose a solution that is based on an extension of ICD-10-PCS with a subset of ICD-9-Proc codes in order to be able to work with legacy data (ICD-9-Proc) together with current data (ICD-10-

* Corresponding authors.
E-mail addresses: ghernandez@informed.dia.fi.upm.es (G. Hernandez-Ibarburu), dperez@infomed.dia.fi.upm.es (D. Perez-Rey).

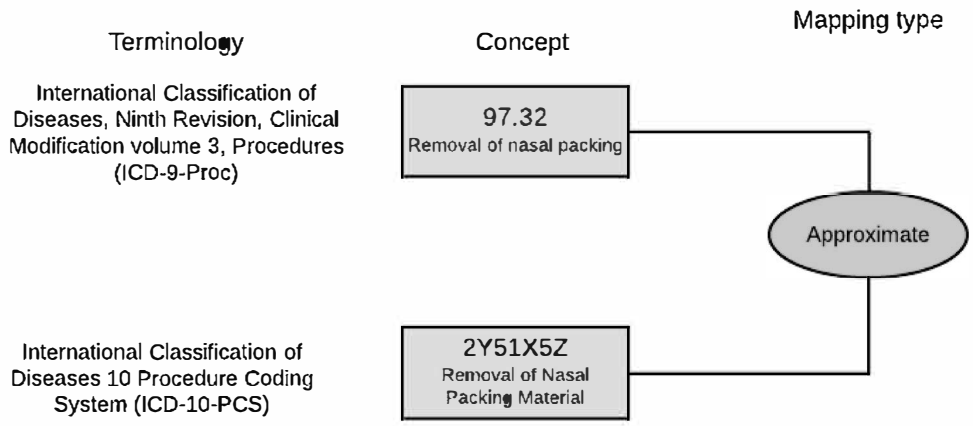

Fig. 1. Example of a concept mapping.

PCS) minimizing data accuracy issues. The extension was developed to be used as part of the reference terminology of the InSite[1] network, an initiative for secondary use of EHR data in Clinical Research.

2. Background

For 30 years, ICD-9-Proc has been the main coding system for health interventions. However, as it is based on the medical knowledge of the mid-1970s, it has become rather obsolete. Despite periodical updates, ICD-9-Proc has fallen behind the evolution of current medical knowledge. An example of a common issue found in ICD-9-Proc is the classification of concepts without enough granularity (e.g., certain laparoscopies are classified along with open surgeries) [5], which might be required for proper reimbursement payments and for medical reporting with sufficient accuracy according to physicians judgement [6].

ICD-9-Proc has around 3882 billable medical procedure codes (around 4700 if node concepts are also considered). A leaf ICD-9-Proc code consists of: two digits, a dot and one or two digits more. The codes 55.61 (Renal autotransplantation) and 36.2 (Heart revascularization by arterial implant) are examples of ICD-9-Proc leaf codes. ICD-9-Proc node codes represent more generic concepts than leafs, e.g. 55.6 (Transplant of kidney) and 36 (Operations on vessels of heart) can be seen as "parents" of leaf nodes which extend their meaning. The top node codes in ICD-9-Proc are numeric ranges, for instance, 55–59 (OPERATIONS ON THE URINARY SYSTEM) and 35–39 (OPERATIONS ON THE CARDIOVASCULAR SYSTEM), they only have the information of the body system where the procedures was carried out.

ICD Tenth Revision, includes several code sets of different knowledge domains. ICD-10-PCS is the set for medical procedure codification. It involves a radical change of approach compared to ICD-9-Proc. The maximum of 4-digits ICD-9-Proc structure of codes is replace by ICD-10-PCS codes that can have up to seven alphanumeric characters. Each one of these characters corresponds to a subsection of a descriptive axis such as anatomy or surgical approach (Fig. 2). This design provides ICD-10-PCS with a big flexibility, allowing easy expansion of the terminology, and much further granularity than ICD-9-Proc. Thus, ICD-10-PCS allows the expression of the diverse and continuously changing world of medical procedures. After some initial training, it also provides an orderly way for humans to find codes [7].

The new main features provided by ICD-10-PCS include granularity, completeness, expandability, different codification for substantial different procedures, intention to avoid "not otherwise specified" (NOS) and "not elsewhere classified" (NEC) concepts and a multiaxial hierarchy [2]. With the tenth version, procedures number of concepts goes to 72,000 (180,000 if nodes concepts are also considered).

However, there are some opponents to the implementation of the new ICD-10-PCS. They claim that, in practice, they don't understand the benefits of this new coding system, at the same time that its implementation requirements are highly expensive [8]. There is also a concern about the training length needed for coders of this terminology, both in the knowledge of the internal structure of ICD-10-PCS and of the detailed anatomy involved in the coded procedures [9]. Many studies have concluded that, at least in the beginning of its implementation, there will be a loss of coding productivity (around 30 and 50 percent) [10] and an increase of coding inaccuracies [11].

With each update of ICD-10-PCS[2], public domain code reference mappings between ICD-9-Proc and ICD-10-PCS, also called General Equivalence Mappings (GEMs), are also provided by the Centers for Medicare and Medicaid Services (CMS) (one for mapping ICD-9-Proc codes to ICD-10-PCS and another one for mapping ICD-10-PCS codes to ICD-9) [12]. However, because the two terminology systems are very different, the documentation accompanying them, and some research made in these mappings, highlights that translating between them most of the time means offering a range of possible options, among which, one must be chosen depending on additional contextual information, rather than offering a semantically equivalent code in the other set [10].

Several research initiatives have used the GEM files to jointly exploit datasets coded with ICD-9 and ICD-10 [13–15]. However, De [16] describes how GEMs should be used to generate a successful mapping. The author claims the GEMs files should be used as guidelines to generate a customized mapping for each institution. He also states that, to create an accurate mapping, both ICD-9-Proc to ICD-10-PCS and ICD-10-PCS to ICD-9-Proc mappings should be considered. Utter et al. [17] published a deep analysis of the GEM accuracy, realizing that certain relationships provided by the GEMs should be revised for each particular case. Out of 212 diagnosis code sets and 64 procedures codes sets, they had to manually edit 95 of them.

To overcome the difficulties of the transition Boyd et al. [18] worked to develop some tools that aims to help healthcare institutions to set less error-prone mappings while showing the complexity of such task for each of the concepts.

In view of the issues of the mapping from ICD-9-Proc procedures to ICD-10-PCS, it was decided that the creation of a customized extension

[1] https://www.insiteplatform.com/

[2] In this paper the terminology version used is the revision of 2017

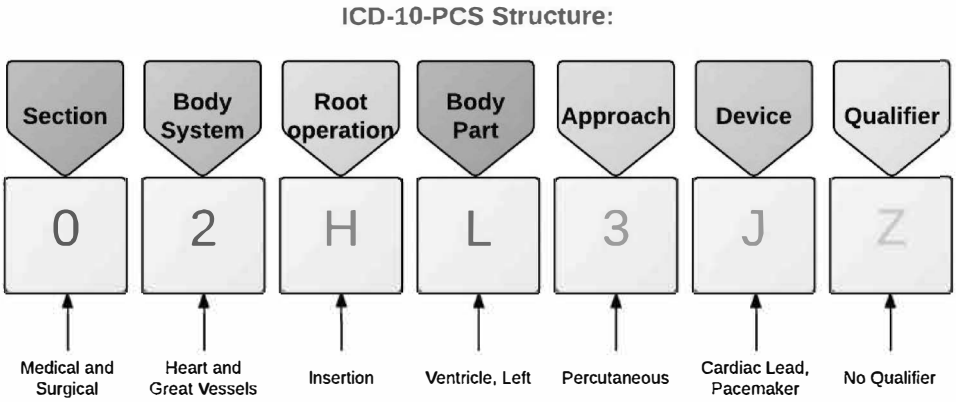

Fig. 2. ICD-10-PCS code structure.

of ICD-10-PCS with ICD-9-Proc codes, using the GEMs relationships to place the new codes in the hierarchy, would enable the integration of both code systems data.

3. Materials and methods

3.1. Incompatibilities between ICD-9-Proc and ICD-10-PCS

Firstly, we present in this paper an analysis of the differences between ICD-9-Proc and ICD-10-PCS terminologies, in order to find the biggest issues of the migration from ICD-9-Proc to ICD-10-PCS.

Most European hospitals use ICD codes to store and report their data about diagnosis and medical procedures. The issue to be solved is that some data in these hospitals are stored in ICD-9-Proc and other in ICD-10. The goal is to achieve the possibility of using these datasets together.

One of the main incompatibilities is due to the large granularity differences between two hierarchy levels of ICD-10-PCS, which does not occur in ICD-9-Proc. In order to be coherent with its established multiaxial hierarchy, some ICD-10-PCS concepts include a greater specificity compared to their hierarchical parents, leading to the need for more general concepts placed in the hierarchy, between the concepts and their parents. For instance, the ICD-10-PCS concept 0SR (Medical and Surgical > Lower Joints > Replacement) has, among its direct descendants, the concepts 0SRC (Medical and Surgical > Lower Joints > Replacement > Knee Joint, Right) and 0SRD (Medical and Surgical > Lower Joints > Replacement > Knee Joint, Left), concepts that include both the body part and the laterality specification. However, how would we map a legacy system with 81.54 (Total knee replacement) data without laterality of the knee replacement stored? It would be necessary to add a new code with only the body part information, descendant of 0SR and parent of 0SRC and 0SRD, to map the ICD-9-Proc concept 81.54, otherwise a patient with one knee replacement can be mistakenly considered to have two knee replacements (one for each side) (see Fig. 3). This laterality issue happens with several other concepts and there are other similar issues (Table 1). This incompatibility is one of the reasons why in the GEMs, provided by the Centers for Medicare and Medicaid Studies (CMS), one ICD-9-Proc concept is mapped to many ICD-10-PCS concepts (sometimes even hundreds!). An increment of granularity is often observed when translating an ICD-9-Proc concept to ICD-10-PCS. However, there are a small number of cases where the opposite occur, the information depth of concepts is reduced in ICD-10-PCS. This is the case of the 37.53 (Replacement or repair of thoracic unit of (total) replacement heart system) and 37.54 (Replacement or repair of other implantable component of (total) replacement heart system) ICD-9-Proc concepts, which are both mapped to 02WA0JZ (Revision of Synthetic Substitute in Heart, Open Approach) by the GEMs (both directions). When querying for the ICD-10-PCS concept using the proposed solution, information related to "type of substitution" might be lost. But in the context of reusing EHR data for research, the mapping is considered valid, since all the semantic information of 02WA0JZ is included in both ICD-9-Proc concepts.

Another important issue is the information in some ICD-9-Proc concepts that is not included in any ICD-10-PCS axis. Some surgical procedures need further information than the provided by ICD-10 PCS axes. In these cases, one of the terminologies specifies information that is not included in the other one and vice versa. For instance, the ICD-9-Proc code 60.21 (Transurethral (ultrasound) guided laser induced prostatectomy (TULIP)) is mapped to 0V507ZZ (Destruction of Prostate, Via Natural or Artificial Opening) and 0V508ZZ (Destruction of Prostate, Via Natural or Artificial Opening Endoscopic) because there is not an axis the information of the guidance imaging technique or the type of destruction procedure.

The ambiguity of ICD-9-Proc concepts meaning in contrast to the specific meaning of ICD-10-PCS concepts is another important source of incompatibility. Many ICD-9-Proc codes are used to group many concepts not coded elsewhere. Some examples of these codes are 38.60 (Other excision of vessel, unspecified site), 13.90 (Operation on lens, Not Elsewhere Classified), 79.90 (Unspecified operation on bone injury, Unspecified site) and 37.99 (Other operations on heart and pericardium). Besides being mapped to a large number of ICD-10 PCS codes (see Table 3), they usually have a very different mapping in each mapping direction (ICD-9-Proc to ICD-10-PCS and ICD-10-PCS to ICD-9-Proc) Table 3.

In many cases, an essential part of the ICD-9-Proc concept information is specified in the last axes of ICD-10-PCS, while it is rather unspecific regarding the information in the first axes. For example, the ICD-9-Proc concept 86.66 (Homograft to skin) specifies in detail the type of tissue graft included (6th axis, Device) but it is at the same time very ambiguous in the body part where the tissue replacement is carried out (4th axis, Body Part), thus, this code is mapped to 40 ICD-10-PCS codes.

All these incompatibilities lead to many semantic incoherences when using GEMs to translate ICD-9-Proc codes to ICD-10-PCS, especially for purposes of aggregation analyses.

3.2. Sources of concepts relationships

The relationships to place the extension codes in ICD-10-PCS hierarchy were extracted from the GEMs (only equality and close match relationships are considered).

Some research made on these mappings has found that, in the mapping from ICD-9-Proc to ICD-10-PCS, the 97.9% of the match type is approximate and just the 0.1% is exact [7]. These results may be explained by the greater specificity of ICD-10-PCS versus ICD-9-Proc, which means that, in these mappings, a single code of ICD-9-Proc is typically linked to multiple ICD-10-PCS codes (all codes that include the

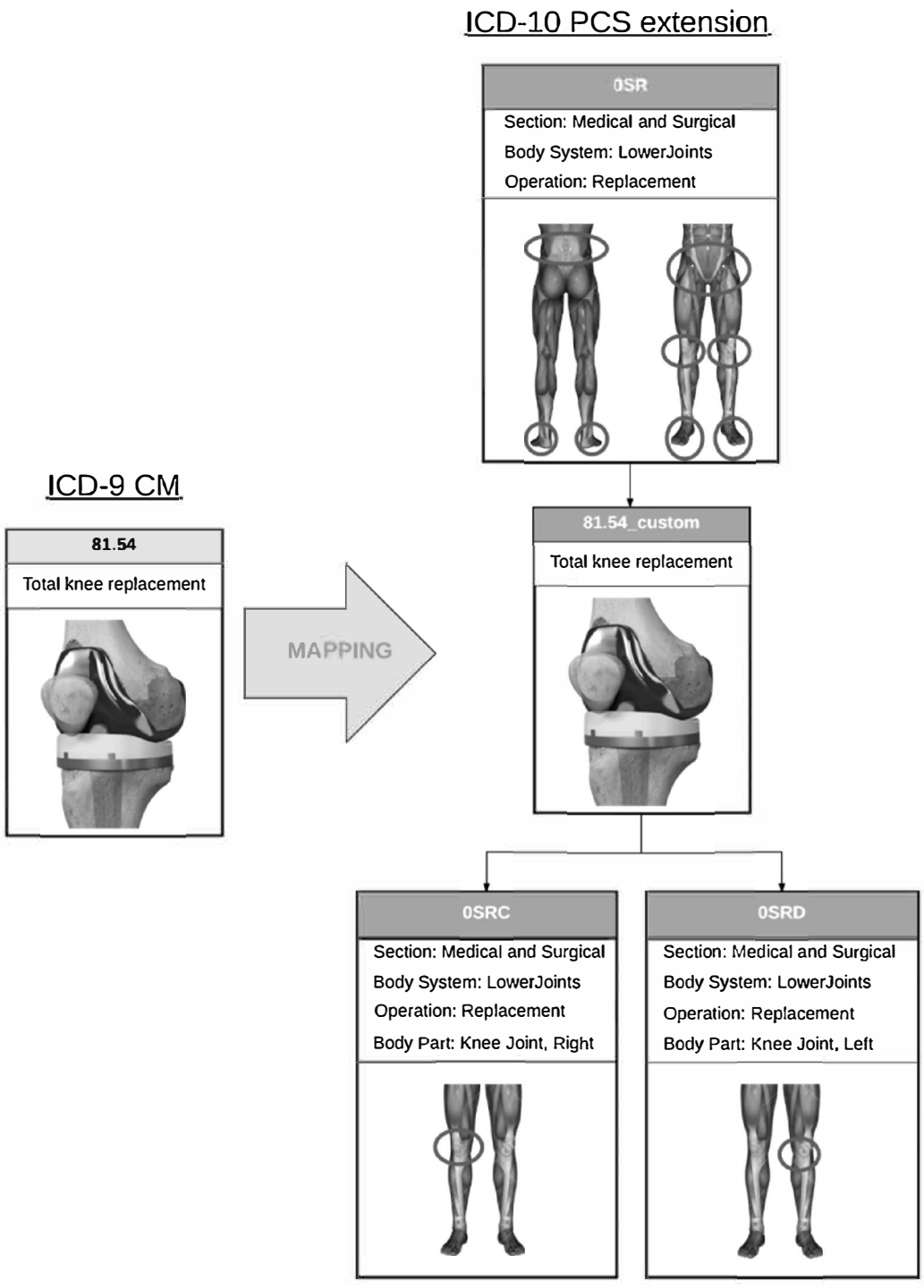

Fig. 3. Example of incompatibility between ICD-9-Proc and ICD-10-PCS and the solution proposed.

meaning of the ICD-9-Proc concept)

3.3. Grouping ICD-9-Proc codes

To select the concepts that could be added automatically to the extension with GEMs relationships, ICD-9-Proc concepts were analyzed and classified into five groups based on their GEMs maps (Fig. 4).

Consider P_9 the set of ICD-9-Proc concepts, P_{10} the set of ICD-10-PCS concepts, S_9 the set of all possible combinations without replacement of P_9 elements and S_{10} the set of all possible combinations without replacement of P_{10} elements. Consider also the function F: $P_9 \rightarrow S_{10}$, which relates, with an "equality" or "approximate" relationship, each ICD-9-Proc concept with the ICD-10-PCS concepts to which it is mapped to according to the GEMs in the forward mapping (ICD-9-Proc to ICD-10-PCS); the function B: $P_{10} \rightarrow S_9$, which relates, with an "equality" or "approximate" relationship, each ICD-10-PCS concept with the ICD-9-Proc concepts to which it is mapped to in the backward mapping of the GEMs (ICD-10-PCS to ICD-9-Proc); and the function H: $P_{10} \rightarrow S_{10}$, which relates each ICD-10-PCS codes to their hierarchical descendants.

Group 1 ($G1$) is defined as the group of all ICD-9-Proc concepts that match only with one ICD-10-PCS concept in both mapping direction ICD-9-Proc to ICD-10-PCS and ICD-10-PCS to ICD-9-Proc. The mapping relationships can be "equality" or "approximate" relations. This group of concepts was decided to be left out of the extension because their

Table 1
Examples of incompatibilities between ICD-9-Proc and ICD-10-PCS.

ICD-10-PCS concept	ICD-10-PCS descendants	ICD-9 concept in-between
0FQ Medical and Surgical > Hepatobiliary System and Pancreas > Repair	0FQ5 ... > Hepatic Duct, Right 0FQ6 ... > Hepatic Duct, Left	51.79 Repair of other bile ducts
0CQ Medical and Surgical > Mouth and Throat > Repair	0CQ0 ... > Upper Lip 0CQ1 ... > Lower Lip	27.51 Suture of laceration of lip
02Q Medical and Surgical > Heart and Great Vessels > Repair	02QW ... > Thoracic Aorta, Descending 02QX ... > Thoracic Aorta, Ascending/Arch	39.54 Re-entry operation (aorta)

one-to-one maps don't cause any semantic inaccuracies.

$$G1 = \{p \varepsilon P_9 \mid \exists\, !q \varepsilon P_{10}\colon\ F(p) = \{q\} \bigwedge p \varepsilon B(q)\}$$

Group 2 ($G2$) comprehends the ICD-9-Proc concept that are related, with an approximate relation, to several ICD-10-PCS concepts in the mapping direction ICD-9-Proc to ICD-10-PCS, at the same time that all those ICD-10-PCS concepts are also related to them in the mapping direction ICD-10-PCS to ICD-9-Proc. After some investigation, it was concluded that in these cases, the ICD-10-PCS concepts tend to be further specifications of the ICD-9-Proc concept.

$$G2 = \{p \varepsilon (P_9 \setminus G1) \mid \forall\ q \varepsilon F(p)\colon\ p \varepsilon B(q)\}$$

Two subgroups can be identified inside $G2$. Group 2.1 includes the group of ICD-10-PCS concepts related to the ICD-9-Proc concept is equal to the group of all existing descendants of an existing ICD-10-PCS node concept. This means that the ICD-9-Proc concept is a synonym of that ICD-10-PCS node.

$$G2.1 = \{p \varepsilon G2 \mid \exists\ q \varepsilon P_{10}\colon F(p) = H(q)\}$$

Group 2.2 comprehends the set of ICD-9 concepts of $G2$ that do not satisfy the condition of Group 2.1.

$$G2.2 = G2 \setminus G2.1$$

For instance, the Group 2 ICD-9-Proc concept 23.09 (Extraction of other tooth) is mapped to all the descendants of 0CD (Medical and Surgical > Mouth and Throat > Extraction) that refer to extraction of tooth but not to the ones that refer to extraction of vocal cord. For this reason, it is classified in G2.2 and not in G2.1.

Group 3 ($G3$) is a group that includes all the ICD-9-Proc concepts that have an "approximate" relation in the ICD-9-Proc to ICD-10-PCS direction of the GEMs but they are not present in the ICD-10-PCS to ICD-9-Proc direction. No reason was found for the absence of these concepts in the ICD-10 PCS to ICD-9-Proc mapping. Even so, as in Group 2, the ICD-10-PCS concepts are further specifications of the ICD-9-Proc concept so they were also added to the extension.

$$G3 = \{p \varepsilon P_9 \mid \exists\ q \varepsilon P_{10}\colon q \varepsilon F(p) \bigwedge B(q) = \emptyset\}$$

Analogously to $G2$, $G3$ can be further divided into two subgroups. Group 3.1 includes the ICD-9-Proc concepts in Group 3 that satisfy the condition described in Group 2.1 explanation.

$$G3.1 = \{p \varepsilon G3 \mid \exists\ q \varepsilon P_{10}\colon F(p) = H(q)\}$$

Group 3.2 comprehends ICD-9-Proc concepts in Group 3 that do not

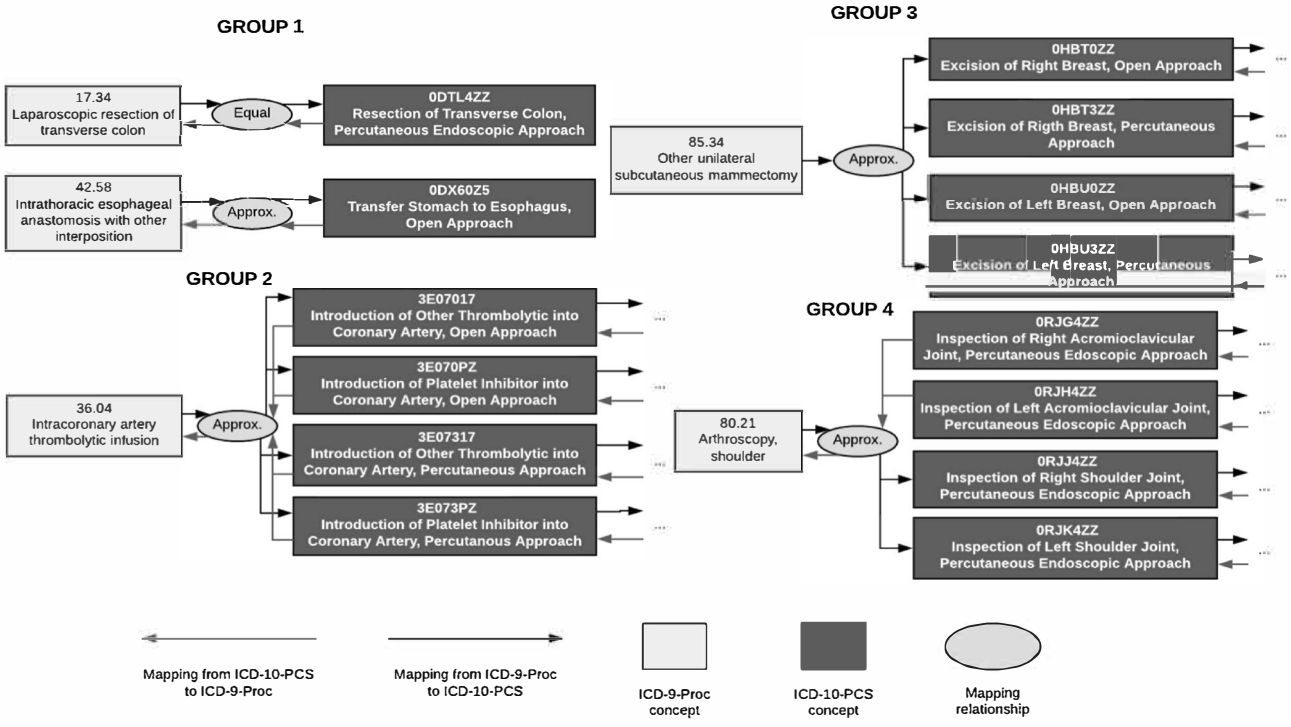

Fig. 4. Classification of ICD-9-Proc concepts based on their GEMs relationships.

satisfy the condition of Group 3.1.

$G3.2 = G3 \setminus G3.1$

Group 4 ($G4$) is composed of all ICD-9-Proc concepts that have "approximate" relations in both ICD-9-Proc to ICD-10-PCS and ICD-10-PCS to ICD-9-Proc mappings, but the relationships in one direction are not the same as the relationships in the other one. No rule was found to be generally applied to this group of whether these concepts should be added or not to the extension, or what mapping direction to choose in case they are added. Sometimes one of the two mapping directions has mapping relationships that are semantically incorrect; sometimes it happens with the other direction, in other cases, one mapping direction does not have all the relationships that should be present, and sometimes the mapping relationships are to completely different concepts between both directions. Aiming to avoid inaccuracies, it was decided that these concepts would be at first left out of the extension, and they would be added later as they are needed, choosing then the mapping direction to use for those concepts.

$$G4 = \{p \varepsilon P_9 \mid \exists\, q \varepsilon P_{10} \colon (\neg\, q \varepsilon F(p) \bigwedge p \varepsilon B(q))(\neg\, p \varepsilon B(q) \bigwedge q \varepsilon F(p))\}$$

Finally, Group 5 ($G5$) consists of all the ICD-9-Proc concepts that don't have an "approximate" or "equality" relationship in any of the two GEMs directions.

$G5 = P_9 \setminus G1 \setminus G2 \setminus G3 \setminus G4$

3.4. Insertion of ICD-9-Proc concepts into ICD-10-PCS hierarchy

Once the ICD-9-Proc concepts to be added to ICD-10-PCS hierarchy were selected, along with their mapping relationships (for Group 3 concepts).

These concepts were set as parents of the ICD-10-PCS concepts to which they were mapped to according the selected relations of the GEMS, as they extend their meaning.

For the settlement of the ICD-10-PCS parent, it was search among all the extension concept ICD-10-PCS children their nearest common ancestor. This ancestor was set as the ICD-9-Proc concept parent as long as it was below the second hierarchy level; otherwise, several ICD-9-Proc codes from the same concept were added and each one had one of the possible third-hierarchical level concepts established as its single parent.

For the correct implementation of the terminology extension it would be advisable to use ontology management software, as it would ease tasks like the connection of concepts or the finding of the ICD-9-Proc concept ancestor.

3.5. Adaptation of mapping

Once the extension was implemented, the mapping of ICD-9-Proc to ICD-10-PCS (with the extension) was adapted. ICD-9-Proc codes that were mapped to just one ICD-10-PCS code (Group 1) were kept mapped as before. ICD-9-Proc codes related to the same set of codes of an ICD-10-PCS node's children (Group 2.1 and 3.1), were mapped to that ICD-10-PCS code. ICD-9-Proc codes that were selected to be introduced automatically in the extension (Group 2.2 and 3.2), were mapped to their homologous extension code. ICD-9-Proc codes that had different mapping relationships in both directions (Group 4) were removed from the mapping. These codes can be added after a domain expert revises manually their mapping relationships for each direction and choose the most suitable ones. Finally, ICD-9-Proc codes that weren't mapped before (Group 5) remained not mapped.

In this step, setting the mapping for Group 2.1 and Group 3.1 concepts could also make use of available ontology management programs that have already implemented the recursive methods needed to navigate through the terminology structure.

Table 2
Number of concepts classified in each group.

Number of ICD-9-Proc concepts included in the extension	Group 2.2	1449
	Group 3.2	565
	Total	2014
Number of ICD-9-Proc concepts mapped to an original ICD-10-PCS concept	Group 1	401
	Group 2.1	164
	Group 3.1	103
	Total	668
Number of ICD-9-Proc concepts ignored	Group 4	856
	Group 5	344
	Total	1200

4. Results

Tables 2 and shows the number of concepts of ICD-9-Proc in each group of the classification, along with the number of concepts that were in the end selected to be added automatically to the ICD-10-PCS extension based on their maps in the GEM files. The 2014 selected concepts that were added to the ICD-10-PCS hierarchy were set as parents of the ICD-10 PCS concepts to which they were mapped to, as these last ones include their meaning and further specifications.

When the ICD-10-PCS concepts children of an extension concept came from different third hierarchy level branches, the selected concept was added one time per branch, as result, a total number of 3052 concepts were added to the ICD-10-PCS extension. The distribution of this extension concepts was analyzed based on the hierarchical level where they were added (Fig. 5) and based on their upper branches of the hierarchical level (Fig. 6). The results show that the majority of the extension concepts were placed under the third hierarchical level of ICD-10-PCS of the Medical and Surgical branch.

The mapping adaptation to the extension (Fig. 7) involves an initial decrease of the number of ICD-9-Proc concepts mapped (24.15%), however, the concepts that are kept unmapped could be added later once they are revised and the appropriate mapping relationships were chosen. The number of ICD-10-PCS (and extension) codes to which the ICD-9-Proc are mapped is also significantly reduced with this adaptation (-92.43%) but, among these maps, there is a sensitive increase of the number of one-to-one maps.

Besides of ICD-9-Proc mapping, the extension codes were also useful to perform mappings of other procedures terminologies such as KVÅ[3] and NCSP[4], with a structure compatible with ICD-9-Proc. Fig. 8.

shows the percentage of extension codes used in such mappings compared to the use of ICD-10-PCS original codes.

As summary, the results of ICD-10-PCS extension with ICD-9-Proc codes by using GEMs relationships showed that the 51.88% (2014/3882) of ICD-9-Proc concepts could be inserted in ICD-10-PCS, 6.88% (267/3882) could be mapped to one ICD-10-PCS node instead of many leafs, as the GEMs suggested, and 10.32% could be kept mapped to the ICD-10-PCS concept that the GEMs proposed. This means that 69.01% (2,682/3882) of ICD-9-Proc concepts could be used together with ICD-10-PCS data in a far more reliable way than by directly using the GEMs. Most of the extension's impact would be noticed in the ICD-10-PCS branch "*Medical and Surgical*", providing further grouping in ICD-10-PCS hierarchical level 3, "Body part" axis.

A manual inspection of the resulting terminology extension and mapping was performed to identify potential limitations of the proposed approach. It was assessed that inserted ICD-9-Proc concepts from Group 2.2 and 3.2 are set as children of ICD-10-PCS concepts that include their meaning. A similar process was performed for new mappings of Group 2.1 and 3.1 concepts. The main limitation is related to ICD-9-Proc concepts that were included in Group 4. Although certain concepts from this group could be mapped and inserted in an extension

[3] http://icd.internetmedicin.se/kva/

[4] http://www.nordclass.se/ncsp_e.htm

Table 3
All maps to ICD-9 code 79.90 (Unspecified operation on bone injury, unspecified site) according to GEMs. The tables shows the combination of characters that represent all ICD-10-PCS codes mapped to this ICD-9-Proc code, 153 codes in total.

Section	Body System	Operation	Body Part	Approach	Device	Qualifier
0 (Medical and Surgical)	P (Upper bones)	Q (Repair)	0 (Sternum)	0 (Open Approach)	Z (No device)	Z (No qualifier)
			1 (1 to 2 Ribs)			
			2 (3 or More Ribs)			
			3 (Cervical Vertebra)			
			4 (Thoracic Vertebra)			
			5 (Right Scapula)			
			6 (Left Scapula)			
			7 (Right Glenoid Cavity)			
			8 (Left Glenoid Cavity)	3 (Percutaneous Approach)		
			9 (Right Clavicle)			
			B (Left Clavicle)			
			C (Right Humeral Head)			
			D (Left Humeral Head)			
			F (Right Humeral Shaft)			
			G (Left Humeral Shaft)			
			H (Right Radius)			
			J (Left Radius)			
			K (Right Ulna)	4 (Percutaneous Endoscopic Approach)		
			L (Left Ulna)			
			M (Right Carpal)			
			N (Left Carpal)			
			P (Right Metacarpal)			
			Q (Left Metacarpal)			
			T (Right Finger Phalanx)			
			V (Left Finger Phalanx)			
			W (Lumbar Vertebra)			
	Q (Lower Bones)	Q (Repair)	1 (Sacrum)	0 (Open Approach)		
			2 (Right Pelvic Bone)			
			3 (Left Pelvic Bone)			
			4 (Right Acetabulum)			
			5 (Left Acetabulum)			
			6 (Right Upper Femur)			
			7 (Left Upper Femur)			
			8 (Right Femoral Shaft)	3 (Percutaneous Approach)		
			9 (Left Femoral Shaft)			
			B (Right Lower Femur)			
			C (Left Lower Femur)			
			D (Right Patella)			
			F (Left Patella)			
			G (Right Tibia)			
			H (Left Tibia)			
			J (Right Fibula)			
			K (Left Fibula)	4 (Percutaneous Endoscopic Approach)		
			L (Right Tarsal)			
			M (Left Tarsal)			
			N (Right Metatarsal)			
			P (Left Metatarsal)			
			Q (Right Toe Phalanx)			
			R (Left Toe Phalanx)			
			S (Coccyx)			

of ICD-9-Proc, an automatic process in this case is not feasible. Table 4 shows a case of a Group 4 concept that should not be mapped because none of the GEMs relations seems to be suitable. However, Table 5 shows two cases where Group 4 concepts could be inserted into the ICD-10-PCS extension, once the correct relations direction is selected. The best direction is usually the most inclusive one, but there are many exceptions to this rule and a manual examination is required in most of the cases.

The extended ICD-10-PCS and the new mapping have been successfully used by InSite Platform to map ICD-9-Proc data. InSite, the largest European Clinical Data Network, has integrate hospitals data from more than 50 hospitals in 10 European countries, many of which provide legacy data coded in ICD-9-Proc and fresh data coded in ICD-10-PCS.

5. Discussion

The proposed ICD-10-PCS extension and the new mapping rules from ICD-9-Proc to such extension are intended to be used as a solution to medical procedure legacy data integration for clinical research purposes.

As stated by Boyd et al. [18], this work shows the high complexity of translating ICD-9-Proc data to ICD-10-PCS. Both approaches aims to solve the issue for two different context, while *Boyd et al.* provide healthcare data coders and researchers with a tool that assist in refining the GEMs mappings, this work describes a methodology that allows a translation of legacy data to an extended ICD-10-PCS terminology in which both ICD-9-Proc and ICD-10-PCS data together at the same time. The automatic mapping explained in this paper also tries to provide a reliable enough mapping when there are enough resources to develop a manual one. However, the user must be aware of the possible errors derived from this automation. Whenever it is possible, the mapping strategy followed Utter et al. [17] is considered to be the best one. Refining the GEMs with clinicians opinion is certainly bound to increase the quality of the resulting mapping, but such valuable resource may not always be available.

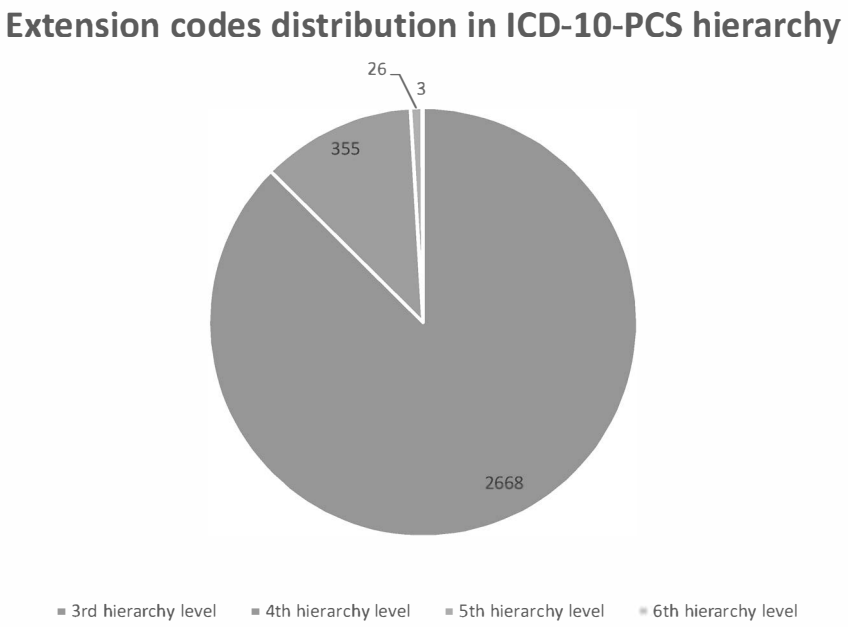

Fig. 5. Extension codes distribution based on the hierarchy level of their ICD-10-PCS parents.

Within the scope of the InSite network, the main purpose of this work is to facilitate the construction of patient cohorts for research analysis. The scope of this article is not to provide a solution for translating ICD-9-Proc billable concepts to ICD-10-PCS billable concepts. As stated in the literature, this task is hardly possible without more information than just the ICD-9-Proc codes of the data.

Another issue that might concern the reader is the maintainability of the extension to keep it updated with the official ICD releases. The new ICD-10-PCS and GEMs releases might change the composition of the stablished ICD-9-Proc groups, so the tracking of ICD-9-Proc group belonging should be revised for each new version, although it is just a subset of 2014 concepts from the more than 180,000 (including leaf and node concepts) that are included in ICD. By following the proposed solution, such maintenance can be performed for future versions of ICD-10-Proc. There are no other maintenance issues as ICD-9-Proc is no longer updated.

6. Conclusions

Common access to diagnosis and procedure coded with ICD-9 and ICD-10 is a challenge for most clinical institutions reusing legacy data. Nowadays, there is a lack of methods to efficiently exploit ICD-9-Proc and ICD-10-PCS data together. The GEMs is a useful resource to migrate coding processes, but they are not intended to solve a common endpoint, since they do not provide a 1-to-1 mapping for every concept of both terminologies. However, they can be used as a start point to find where ICD-9-Proc codes should be place in ICD-10-PCS hierarchy to get an extended terminology that could be used for aggregation purposes.

One of the features of the GEMs that limit their mapping quality is that they only specify mappings from ICD-9-Proc leaf concepts to ICD-10-PCS leaf concepts, and vice versa. Actually, as ICD-10-PCS frequently includes more granularity than ICD-9-Proc, it would often be more useful and accurate to have mappings between ICD-10-PCS node concepts and ICD-9-Proc leaf concepts. With this approach, the number

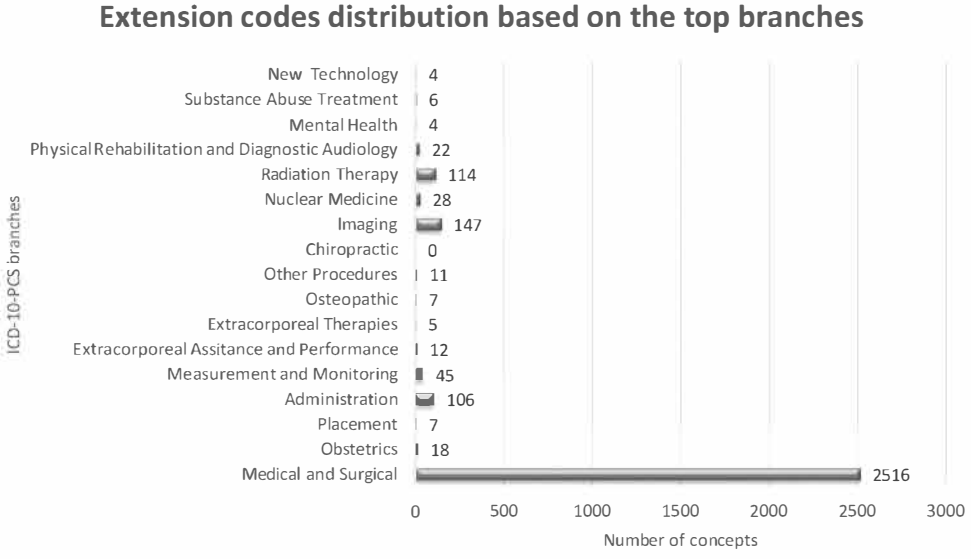

Fig. 6. Extension codes distribution based on their branch in which they were placed.

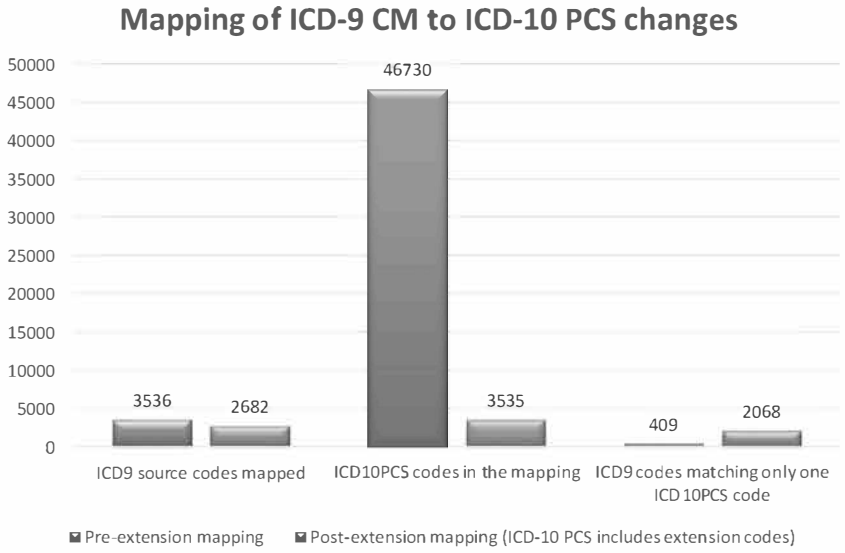

Fig. 7. Changes in the mapping from ICD-9 after the extension.

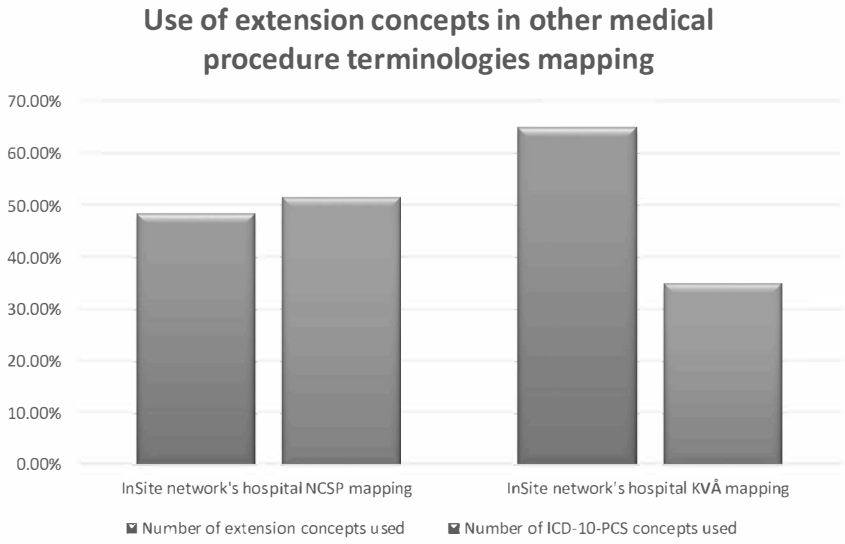

Fig. 8. Use of the extension concepts in mappings from other local hospital's procedure terminologies.

Table 4
Example of G4 ICD-9-Proc concept which is better as not mapped/inserted. "F" relation direction means the GEMs forward direction, and "B" means the backward one.

Code	Description	Relation direction	ICD-10-PCS mapping code	ICD-10-PCS Description
54.12	Reopening of recent laparotomy site	F	0W3G0ZZ	Control Bleeding in Peritoneal Cavity, Open Approach
			0W3H0ZZ	Control Bleeding in Retroperitoneum, Open Approach
			0W3P0ZZ	Control Bleeding in Gastrointestinal Tract, Open Approach
			0WJG0ZZ	Inspection of Peritoneal Cavity, Open Approach
			0WJH0ZZ	Inspection of Retroperitoneum, Open Approach
			0WJJ0ZZ	Inspection of Pelvic Cavity, Open Approach
		B	0W3G0ZZ	Control Bleeding in Peritoneal Cavity, Open Approach
			0W3H0ZZ	Control Bleeding in Retroperitoneum, Open Approach
			0W3P0ZZ	Control Bleeding in Gastrointestinal Tract, Open Approach

of 1-to-1 mappings would be increased and the result mapping would be improved. The GEMs can be used to identify these new mappings and also where to insert (set the right direct parents and direct children) other ICD-9-Proc concepts in ICD-10-PCS. The final aim is to facilitate the integration of the majority of ICD-9-Proc data in ICD-10-PCS.

After the development this extension, a future line of work would be to apply a similar approach to ICD-10-CM in order to solve the issue of translating data from ICD-9-CM diagnosis codes (ICD-9-Diag). For this translation, the GEMs are also available, but again they fail to provide a reliable mapping between both terminologies. A similar approach could be applied, and the results could be even better, as ICD-10-CM and ICD-9-Diag have less differences than the procedure area.

Table 5
Example of G4 ICD-9-Proc concepts that could be inserted. "F" relation direction means the GEMs forward direction, and "B" means the backward one.

Code	Description	Relation direction	ICD-10-PCS mapping code	ICD-10-PCS Description
88.92	Magnetic resonance imaging of chest and myocardium	F	B236Y0Z	Magnetic Resonance Imaging (MRI) of Right and Left Heart using Other Contrast, Unenhanced and Enhanced
			B236YZZ	Magnetic Resonance Imaging (MRI) of Right and Left Heart using Other Contrast
			B236ZZZ	Magnetic Resonance Imaging (MRI) of Right and Left Heart
			BW33Y0Z	Magnetic Resonance Imaging (MRI) of Chest using Other Contrast, Unenhanced and Enhanced
			BW33YZZ	Magnetic Resonance Imaging (MRI) of Chest using Other Contrast
		B (selected)	B236Y0Z	Magnetic Resonance Imaging (MRI) of Right and Left Heart using Other Contrast, Unenhanced and Enhanced
			B236YZZ	Magnetic Resonance Imaging (MRI) of Right and Left Heart using Other Contrast
			B236ZZZ	Magnetic Resonance Imaging (MRI) of Right and Left Heart
			BB3GY0Z	Magnetic Resonance Imaging (MRI) of Lung Apices using Other Contrast, Unenhanced and Enhanced
			BB3GYZZ	Magnetic Resonance Imaging (MRI) of Lung Apices using Other Contrast
			BB3GZZZ	Magnetic Resonance Imaging (MRI) of Lung Apices
			BW33Y0Z	Magnetic Resonance Imaging (MRI) of Chest using Other Contrast, Unenhanced and Enhanced
			BW33YZZ	Magnetic Resonance Imaging (MRI) of Chest using Other Contrast
48.36	[Endoscopic] polypectomy of rectum	F (selected)	0DBP4ZZ	Excision of Rectum, Percutaneous Endoscopic Approach
			0DBP8ZZ	Excision of Rectum, Via Natural or Artificial Opening Endoscopic
		B	0DBP8ZZ	Excision of Rectum, Percutaneous Endoscopic Approach

Declaration of conflicting interests

The authors have no conflicts of interest to declare.

Acknowledgment

This work has been funded by the project Development of tools to integrate clinical trial data. which is a collaboration between InSite Platform and the Biomedical Informatics Group of Universidad Politécnica de Madrid.

References

[1] G. Wade, S.T. Rosenbloom, Experiences mapping a legacy interface terminology to SNOMED CT, BMC Med. Inform. Decis. Mak. 8 (Suppl 1) (2008) S3, Oct.
[2] R. Mullin, A brief history of ICD-10-PCS, J. AHIMA 70 (October 9) (1999) 97–98.
[3] G. Rahmathulla, et al., Migration to the ICD-10 coding system: a primer for spine surgeons (Part 1), Surg. Neurol. Int. 5 (Suppl 3) (2014) S185–S191.
[4] G. Rahmathulla, et al., Implementation and impact of ICD-10 (Part II), Surg. Neurol. Int. 5 (Suppl 3) (2014) S192–S198.
[5] M.C. Libicki, I.T. Brahmakulam, The Costs And Benefits of Moving To The Icd-10 Code Sets, RAND Corporation, Santa Monica, CA, 2005.
[6] C. Benesch, D.M. Witter, A.L. Wilder, P.W. Duncan, G.P. Samsa, D.B. Matchar, Inaccuracy of the International Classification of Diseases (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease, Neurology 49 (January 3) (1997) 660–664.
[7] S.J. Steindel, International classification of diseases, 10th edition, clinical modification and procedure coding system: descriptive overview of the next generation HIPAA code sets, J. Am. Med. Inform. Assoc. 17 (May 3) (2010) 274–282.
[8] J.A. Hirsch, F.J.E. Falco, L. Manchikanti, Necessity and implications of ICD-10: facts and fallacies, Pain Phys. 15 (April 2) (2012) E153.
[9] P. E. Brooks, "Testing ICD-10-PCS," Test. ICD-10-PCS AHIMA Am. Health Inf. Manag. Assoc., May 1998.
[10] S.V. Ross-Davis, Preparing for ICD-10-CM/PCS: One Payer's Experience with General Equivalence Mappings (GEMs)," Perspect. Health Inf. Manag. AHIMA Am, Health Inf. Manag. 9 (January) (2012) Winter.
[11] G.H. Utter, G.L. Cox, P.L. Owens, P.S. Romano, Challenges and opportunities with ICD-10-CM/PCS: implications for surgical research involving administrative data, J. Am. Coll. Surg. 217 (September 3) (2013) 516–526.
[12] S.V. Ross-Davis, Preparing for ICD-10-CM/PCS: one payer's experience with general equivalence mappings (GEMs)," perspect. Health inf. Manag. AHIMA Am, Health Inf. Manag. 9 (January) (2012) Winter.
[13] C. Feudtner, J.A. Feinstein, W. Zhong, M. Hall, D. Dai, Pediatric complex chronic conditions classification system version 2: updated for ICD-10 and complex medical technology dependence and transplantation, BMC Pediatr. 199 (August14) (2014).
[14] J.W. Sun, et al., Adaptation and validation of the combined comorbidity score for ICD-10-CM, Med. Care 55 (Decenber12) (2017) 1046.
[15] A.D. Boyd, et al., Metrics and tools for consistent cohort discovery and financial analyses post-transition to ICD-10-CM, J. Am. Med. Inform. Assoc. 22 (May 3) (2015) 730–737.
[16] S. De, 8 Steps to Success in ICD-10-CM/PCS Mapping: Best Practices to Establish Precise Mapping Between Old and New ICD Code Sets, J. AHIMA 83 (June 6) (2012) 44–49.
[17] G.H. Utter, G.L. Cox, O.O. Atolagbe, P.L. Owens, P.S. Romano, 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 Serv. Res. (May) (2018).
[18] A.D. Boyd, et al., ICD-10 procedure codes produce transition challenges, AMIA Jt. Summits Transl. Sci. Proc. AMIA Jt. Summits Transl. Sci. 2017 (2018) 35–44.