Melissa P. Resnick, Elena Cardillo, Marc Jamoulle, Magdala de Araujo Novaes, Frank S. Shamenek

Towards the semantic annotation and the prevention of the loss of information of second opinion requests from rural Brazilian primary healthcare providers: the Q-codes use case – a work in progress

Abstract

Objectives. To support documentation, various terminologies have been created to assist in this activity. Only a few terminologies cover the General Practice / Family Medicine (GP/FM) domain (*e.g.*, the International Classification of Primary Care - ICPC). As ICPC fails to capture some non-clinical issues (*e.g.*, organizational and managerial aspects), the Q-Codes taxonomy has been developed to extend ICPC, encompassing those contextual professional issues. The aim of this work is to show the value of Q-Codes in preventing loss of information through the semantic annotation of Second Opinion Requests of rural Brazilian primary healthcare providers.

Methods. Question-answer pairs for the years 2010-2012, in the Brazilian-Portuguese language, was obtained from an urban telehealth center. Each selected question was read to determine its semantic meaning, and coded using both the ICPC and Q-Codes classification systems. Based on this meaning, each question was manually assigned between 0 and 5 Q-Codes.

Results. The majority of Q-Code assignments were almost equally split between the Patient's Category", QC (42%) and the "Family Doctor's Issue", QD (37%) domains. Domain QT, that is "Knowledge Management", covered 21% of the assignments, while a single assignment was made to the domain QP, “Patient Issue”. Six of the top 10 Q-Codes assigned belong to the cited QC domain, 3 to the QD, and 1 to the QT domain.

Discussion. Analysis demonstrates that lost information represents age (QC), health prevention (QD), and medical education (QT). Medical education is one of the reasons that the telehealth system was implement, while age and health prevention are important to providing healthcare in Brazil.

Conclusions. Preliminary results show that Q-Codes capture information that otherwise would be lost in the case of using only clinical coding systems such as ICPC.

1. Introduction

Medicine is a broad field with many specialties. To support documentation, storage, and retrieval of information, various terminologies have been created to assist with these activities. These include, but are not limited to, the International Classification of Diseases (ICD) (Lagasse *et al*. 2001), the Medical Subject Headings (MeSH) of the National Library of Medicine (Nelson 2009), and SNOMED-CT nomenclature (Wang *et al*. 2008). The majority of specialized medical domains have proper nomenclatures and classifications, mostly targeted at clinical and specialized medicine (Cornet and de Keizer 2008), thus only partially covering the General Practice / Family Medicine (GP/FM) domain. With the use of primary care related classifications, for example the International Classification of Primary Care (ICPC), it is still difficult for General Practitioners (GPs) or Family Doctors to target and code, in their daily practice, the non-clinical issues or contextual information (*e.g.*, organizational and managerial

aspects) with the available coding systems.

To this aim, during the last few years the Q-Codes taxonomy has been proposed (Jamoulle *et al*. 2017), providing an extension of ICPC concepts for contextual issues, and focusing on the semantic aspects thus constructing a true semantic resource to be used as an indexing system.

In this paper, the feasibility and usefulness of Q-Codes is tested to provide a semantic annotation (using both ICPC-2 for clinical concepts, and Q-Codes for both non-overlapping clinical concepts and non-clinical concepts) of *questions* from Question-Answer pairs of rural Brazilian healthcare providers, as they seek, and receive, second opinions from urban telehealth centers. The aim of this work is to show the value of Q-Codes in preventing loss of information through the semantic annotation of Second Opinion Requests (SOR) of rural Brazilian primary healthcare providers.

2. Background

2.1. International Classification of primary care (ICPC)

The International Classification of Primary Care (ICPC) is a classification initially used for data retrieval in primary care (Boot and Meijman 2010). By 1987, ICPC was introduced by the WONCA International Classification Committee (WICC) (Soler *et al*. 2008; I. Okkes *et al*. 2000), and the second release, including additional inclusion/exclusion notes, was published in 1998. This release is updated online once per year, and the last international update (ICPC-2e-v.6.0) was published in April 2017 (World Organization of Family Doctors (Wonca) and Wonca International Classification Committee (WICC) 2017). GP/FM is very broad in scope, encompassing both clinical and contextual issues (Jamoulle *et al*. 2017). Clinical issues pertain to signs and symptoms, reasons for encounter, processes and diagnoses, which are covered by ICPC (I. Okkes *et al*. 2000).

As noted by the World Health Organization (WHO) (Who 2017), ICPC has:

> a biaxial structure and consists of 17 chapters, each divided into 7 components dealing with symptoms and complaints (comp. 1), diagnostic, screening and preventive procedures (comp. 2), medication, treatment and procedures (comp. 3), test results (comp. 4), administrative (comp. 5), referrals and other reasons for encounter (comp. 6) and diseases (comp. 7).

This classification has been used for structured documentation of episode-oriented care in primary care since the 1980's (Lamberts and Hofmans-Okkes 1996), and is now considered as a de facto standard in Primary Health Care.

However, ICPC only offers a partial solution as it covers only the clinical issues of GP/FM (Boot and Meijman 2010; Soler *et al*. 2008).

Over the years, extensions to ICPC for nutritional advice (van Binsbergen and Drenthen 1999), procedures (I. M. Okkes, Veldhuis, and Lamberts 2002), community

pharmacy (van Mil, Brenninkmeijer, and Tromp 1998), and chiropractic medicine (Testern, Hestbæk, and French 2015; Charity *et al*. 2013) have been developed.

As ICPC fails to capture some non-clinical issues, which are predominately organizational and managerial aspects of GP/FM, a newly developed hierarchical resource, called "Q-Codes", has been developed to extend ICPC encompassing those contextual professional issues. The letter Q was used as it was available for use as a chapter in ICPC.

2.2. Q-Codes

The development of the Q-Codes taxonomy started from the 1987 work of Henk Lamberts, a Dutch professor in General Practice (Amsterdam University), who designed the system for topographic archiving of copies of retrieved articles in the documentation system of the research department of General Practice. It was a simple one level classification with 7 main domain categories. In 2007, MJ, a co-author of ICPC (I. Okkes *et al*. 2000), undertook to revise and develop the Q-Codes as a full extension of ICPC for contextual issues, focusing on the semantic aspects and constructing a true semantic resource to be used as an indexing system for grey literature (Jamoulle *et al*. 2017). The Q-Codes taxonomy consists of 182 terms, distributed among 8 domains (Jamoulle *et al*. 2017), each containing between 2 and 4 levels of granularity, which represents respectively 44 subcategories, 109 sub-subcategories, and 21 sub-sub-subcategories (Jamoulle 2016). The 8 domains include: "Patient's Category", "Family Doctor's Issue", "Medical Ethics", "Planetary Health", "Patient Issue", "Research", "Structure of Practice" and "Knowledge Management" (Jamoulle and Resnick 2016).

The "Patient's Category" domain (QC) represents such concepts as *age*, *gender issues*, and *social issues*. The "Family Doctor's Issue" domain (QD) describes *communication*, *clinical prevention*, and *medico legal issues*. The "Medical Ethics" domain (QE) covers *bioethics*, *professional ethics*, and *infoethics*. The "Planetary Health" domain (QH) deals with such areas as *environmental health*, *biological hazards*, and *nuclear hazards*. The "Patient's Issue" domain (QP) includes *patient safety*, *patient centeredness*, and *quality of healthcare*. The "Research" domain (QR) describes *research methods*, *research tools* and *epidemiology of primary care*. The "Structure of Practice" domain (QS) covers such topics as *primary care setting*, *primary care provider*, and *practice relationship*. Finally, the "Knowledge Management" domain (QT) deals with *teaching*, *training* and *knowledge dissemination* (Jamoulle *et al.* 2017; Jamoulle and Resnick 2016).

Representing non-clinical issues, the Q-Codes taxonomy provides a resource to facilitate access to GP/FM information. The first aim of Q-Codes is its use as an indexing system for grey literature (Jamoulle *et al.* 2017), and e-learning applications.

Constructed on the basis of Semantic web technologies, Q-Codes could be considered as a lightweight ontology ready to be used in the semantic web domain, to be extracted in Web Ontology Language (OWL). We will describe, as a use case for the application of Q-Codes, the Second Opinion Requests (SOR) from rural healthcare settings in the state of Pernambuco, Brazil. One source of information is SOR from rural healthcare teams in Brazil (Resnick *et al*. 2013).

2.3. Second Opinion Requests

Rural healthcare teams (physicians, nurses, lay community health workers) provide basic care to those living in their area (Haddad *et al*. 2015; Sanches *et al*. 2012). Sometimes, however, the healthcare teams need to send their patients to urban areas for a second opinion (consultation) or to see a specialist, often requiring a great deal of travel at considerable costs (Alkmim *et al*. 2012). Lack of expertise amongst health professionals in the primary care sector, unnecessary referrals, and the difficulty of facilitating consultations with specialists led to the development of the Brazilian telehealth program (Joshi *et al*. 2011).

In 1999 the first module of the telehealth platform called HealthNet was set up at Nucleo de Telessaude (NUTES), which is located within the Clinical Hospital of the Federal University of Pernambuco (UFPE) in Recife, Brazil (Barbosa, de A Novaes, and de Vasconcelos 2003). By 2004, HealthNet was implemented and operational (de Araújo Novaes *et al.* 2005). The telehealth service, through HealthNet at NUTES, provides medical second opinions to the healthcare teams in the rural areas of Pernambuco.

When a rural healthcare team needs a second opinion, in order to provide care to a patient, they send their questions through HealthNet to the nurses and physicians at NUTES. The appropriate health professional provides a second opinion or an answer through HealthNet back to the rural healthcare team. These questions and their corresponding answers (question-answer pairs) are collected for data sharing and reuse.

Managers of these telehealth programs need a way to evaluate and plan interventions, which will, in turn, improve access to telehealth services. One way that this can be done is through the use of information classifications, like the Q-Codes. The question-answer pairs from the telehealth service at NUTES will be used for this study.

3. Methodology

A data set containing 5,580 question-answer pairs for the years 2010-2012, in the Brazilian-Portuguese language, was obtained from an urban telehealth center. Webinars and tele-ECG Q/A pairs were eliminated, giving 1,669. Among these, 550 questions (~33% from each of the three years) were randomly selected and deidentified for inclusion into the sample data set.

Each selected question was read by the first author to determine its semantic meaning, and coded using both the ICPC and Q-Codes classification systems. Based on this meaning, some general guidelines and the definitions of individual Q-Codes, each question was manually assigned between 0 and 5 Q-Codes.

When the question provided an age of the patient, the appropriate age group was assigned from the "Patient's Category" domain (QC). When the question pertained to gender issues, such as pregnancy or birth control, it was assigned appropriate concepts from the "Patient's Category" domain (QC). When the question represented a need for information not referring to a specific patient, it was assigned a concept from the "Knowledge Management" domain (QT). Finally, if the question represented disease prevention and multimorbidity, it was assigned the appropriate concepts from the "Family Doctor's Issue" domain (QD).

4. Preliminary results

As of the writing of this paper, 100 (18%) of the 550 questions from the sample data set have been attempted to be semantically annotated with Q-Codes. Out of the 100 attempts, 98 (98%) were successful. Unsuccessful attempts (2%) were due to the lack of semantic meaning in the question (i.e., “If it is altered, forward physician p?”).

For the successfully annotated questions, between 1 and 3 Q-Codes were assigned. Nearly three-fifths (56%) of the questions were assigned 2 Q-Codes; 41% were assigned 1 Q-Code; and 3% were assigned 3 Q-Codes. There were seven instances where the question was assigned at least one Q-Code, while being unable to be coded with any ICPC codes.

A cumulative total of 159 Q-Codes were assigned to the 98 questions. More precisely, 97.5% of these Q-Codes were assigned at the sub-subcategory level, with 1.25% of the Q-Codes being assigned at both the subcategory and the sub-sub-subcategory level.

The vast majority of Q-Code assignments were almost equally split between the QC (42%) and QD (37%) domains. Domain QT covered 21% of the assignments, while a single assignment was made to the domain QP. Four domains were not assigned: QE, QH, QR and QS. Six of the top 10 Q-Codes assigned belong to the Patient's Category domain (QC), 3 to the Doctor’s Issue domain (QD), and 1 to the Knowledge Management domain (QT).

5. Discussion

Of the top 10 assigned Q-Codes, only one category (QC22 "Women's health") overlaps partially with ICPC, and thus, does not provide much additional information about the semantic meaning of the question asked. Additionally, some questions represent a need for information on clinical topics not associated with a specific patient,

which could be used for future cases; "Continuous Medical Education" (QT23) was used for these questions. In fact, one of the reasons for the implementation of the telehealth system in Brazil is to provide medical education (Alkmim *et al*. 2012; Joshi *et al.* 2011). The remaining eight categories from Table 2 provide information that would be lost if a clinical KO system (i.e., ICPC) was used instead of the Q-Codes. Five of the remaining nine Q-Codes represent age groups (QC11, QC12, QC13, QC14, and QC15). Age of the patient is important, allowing the general practitioner to provide the best and most appropriate care to his/her patients. Three of the top 10 assigned Q-Codes deal with prevention (QD41, QD42, and QD43). This is not surprising, as health prevention and promotion is important to providing primary care in Brazil (Alkmim *et al*. 2012).

Table 1: Top 10 Q-Codes assigned

Rank	**Q-Code**	**Title**	**Number of assignments**
1	QC22	Women's health	23
2	QT23	Continuous medical education	20
3	QD43	Tertiary prevention	20
4	QD42	Secondary prevention	10
5	QC14	Adult	8
6	QD41	Primary prevention	5
7	QC15	Elderly	4
8	QC13	Adolescent	4
9	QC12	Child	3
10	QC11	Infant	3

As a Work in Progress, only 100 questions have been manually annotated by the submission deadline.

One limitation can be seen in the use of manual annotation, which reduces the ability to index large data sets, quality and number of the resulting annotations; is time consuming; and requires number of actions, including inter-annotator agreement (Névéol, Islamaj Doğan, and Lu 2011). A second limitation is the present inability to publish the deidentified data set in a publicly available data repository.

6. Conclusion and Future Work

The use of Q-Codes to annotate Second Opinion Requests from rural Brazilian primary healthcare providers has been tested to show its feasibility in facilitating

communications and coding among rural healthcare providers when expressing non-clinical and contextual issues. Results show that Q-Codes add value capturing information that otherwise would be lost if using only clinical systems such as ICPC.

Future work would include the research goal is to annotate all 550 questions contained within the sample data set. Depending upon the availability of resources, it might be desirable to annotate the entire data set of 1.669 questions.

To overcome the limitation of manual annotation, future work could investigate the use of semi-automated annotation methods to assign Q-Codes to large data sets quickly and more efficiently. Other improvements can be the involvement of a second annotator to validate the dataset; and the use of DeCS to index the SOR to test its feasibility.

To overcome the lack of public access to the deidentified dataset, UFPE NUTES has a "plan to do this through a new component in our telehealth platform, but probably available by the end of 2018" (Magdala de Araújo Novaes, e-mail message to author, December 6, 2017).

Acknowledgments

The authors thank Nilma Andrade, Stefan Darmoni, Paula Rejane Beserra Diniz, Julien Grosjean, Ashwin Ittoo, Fernando José Ribeiro Sales, Robert Vander Stichele, and Marc Vanmeerbeek.

References

Alkmim, Maria Beatriz, Renato Minelli Figueira, Milena Soriano Marcolino, Clareci Silva Cardoso, Monica Pena de Abreu, Lemuel Rodrigues Cunha, Daniel Ferreira da Cunha, *et al.* (2012). Improving Patient Access to Specialized Health Care: The Telehealth Network of Minas Gerais, Brazil. *Bulletin of the World Health Organization* 90(5): 373-378. Available at:. https://doi.org/10.2471/BLT.11.099408.

Araújo Novaes, Magdala de, Ana Karina Pinto Barbosa, Kleber Soares de Araújo, Jeane M. Lacerda de A Couto, Glauber Araújo, and Leonardo Sarmento (2005). Experiences on the Use of a Second Opinion Software for the Primary Care. *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, p. 889.

Barbosa, Ana Karina P., Magdala de A Novaes, and Alexandre M. L. de Vasconcelos (2003). A Web Application to Support Telemedicine Services in Brazil. *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, p. 56-60.

Binsbergen, J. J. van, and A. J. Drenthen (1999). ICPC-Code Approach of Nutritional Questions in General Practice: a look at the future. *European Journal of Clinical Nutrition* 53(Suppl 2) (May): S22-24.

Boot, Cécile R. L., and Frans J. Meijman (2010). Classifying Health questions asked by the public using the ICPC-2 Classification and a Taxonomy of Generic Clinical Questions: an empirical exploration of the feasibility. *Health Communication* 25 (2):175-181. Available at: https://doi.org/10.1080/10410230903544969.

Charity, Melanie J., Simon D. French, Kirsty Forsdike, Helena Britt, Barbara Polus, and Jane Gunn (2013). Extending ICPC-2 PLUS Terminology to Develop a Classification System Specific for the Study of Chiropractic Encounters. *Chiropractic & Manual Therapies* 21 (1):4. Available at: https://doi.org/10.1186/2045-709X-21-4.

Cornet, Ronald, and Nicolette de Keizer (2008). Forty Years of SNOMED: a literature review. *BMC Medical Informatics and Decision Making* 8(Suppl 1) (October): S2. Available at: https://doi.org/10.1186/1472-6947-8-S1-S2.

Haddad, Ana Estela, Mary Caroline Skelton-Macedo, Verônica Abdala, Caren Bavaresco, Daniele Mengehel, Camilla Galatti Abdala, and Erno Harzheim (2015). Formative Second Opinion: Qualifying Health Professionals for the Unified Health System through the Brazilian Telehealth Program. *Telemedicine Journal and E-Health: the official journal of the American Telemedicine Association* 21 (2):138 -142. Available at: https://doi.org/10.1089/tmj.2014.0001.

Jamoulle, Marc (2016). The Q-Codes – 3CGP: Core content classification in general practice family medicine. 2016. Available at: http://3cgp.docpatient.net/tabular/.

Jamoulle, Marc, Julien Grosjean, Melissa Resnick, Ashwin Ittoo, Arthur Treuherz, Robert Vander Stichele, Elena Cardillo, Stéfan J. Darmoni, Frank S. Shamenek, and Marc Vanmeerbeek (2017). A Terminology in General Practice/Family Medicine to represent non-clinical aspects for various usages: The Q-Codes. *Studies in Health Technology and Informatics* 235:471-75.

Jamoulle, Marc, and Melissa P. Resnick (2016). *General Practice / Family Medicine Multilingual Terminology – English Version*. Strépy-Bracquegnies, Belgium: Le livre en papier. Available at: http://www.publier-un-livre.com/fr/le-livre-en-papier/349-general-practice-family-medicine-multilingual-terminology-english-version.

Joshi, Ashish, Magdala A. Novaes, Sriram Iyengar, Josiane L. Machiavelli, Jiajie Zhang, Robert Vogler, and Chiehwen E. Hsu (2011). Evaluation of a Tele-Education Programme in Brazil. *Journal of Telemedicine and Telecare* 17 (7):341-345. Available at: https://doi.org/10.1258/jtt.2011.101209.

Lagasse, R, M Desmet, M Jamoulle, G Correa, M Roland, P Hoyois, and Ch De Brouwer (2001). "European Situation of the Routine Medical Data Collection and Their Use for the Health Monitoring (Euro-Med-Data). Final Report 1998/IND/2011 / SI2.107874. Brussels, Belgium: Université Libre de Bruxelles, Ecole de Santé Publique. http://www.ulb.ac.be/esp/emd/Emd_rep.pdf.

Lamberts, H., and I. Hofmans-Okkes (1996). The Core of Computer Based Patient Records in Family Practice: Episodes of Care Classified with ICPC. *International Journal of Bio-Medical Computing* 42 (1-2): 35-41.

Mil, J. W. van, R. Brenninkmeijer, and T. F. Tromp (1998). The ICPC Coding System in Pharmacy: Developing a Subset, ICPC-Ph. *Pharmacy World & Science: PWS* 20(1): 38-42.

Nelson, S. J. (2009). "Medical Terminologies That Work: The Example of MeSH." In *2009 10th International Symposium on Pervasive Systems, Algorithms, and Networks*, 380-384. Available at: https://doi.org/10.1109/I-SPAN.2009.84.

Névéol, Aurélie, Rezarta Islamaj Doğan, and Zhiyong Lu (2011). Semi-Automatic Semantic Annotation of PubMed Queries: a study on quality, efficiency, aatisfaction." *Journal of Biomedical Informatics* 44(2):310–18. https://doi.org/10.1016/j.jbi.2010.11.001.

Okkes, I., M. Jamoulle, H. Lamberts, and N. Bentzen (2000). ICPC-2-E: the electronic version of ICPC-2. Differences from the Printed Version and the Consequences. *Family Practice* 17 (2):101-107.

Okkes, I. M., M. Veldhuis, and H. Lamberts (2002). Severity of Episodes of Care assessed by Family Physicians and Patients: The DUSOI/WONCA as an extension of the International Classification of Primary Care (ICPC). *Family Practice* 19(4):350-356.

Resnick, Melissa P., Filipe Santana, Magdala de Araujo Novaes, Frank S. Shamenek, Lex Frieden, and M. Sriram Iyengar (2013). Representing Second Opinion Requests from Primary Care within the Brazilian Tele-Health Program: International Classification of Primary Care, Second Edition. *Studies in Health Technology and Informatics* 192:1.190.

Sanches, Luiz M. P., Danielle S. Alves, Maria Helena B. M. Lopes, and Magdala A. Novaes (2012). The Practice of Telehealth by Nurses: an experience in primary healthcare in Brazil." *Telemedicine Journal and E-Health: the Official Journal of the American Telemedicine Association* 18 (9):679-683. https://doi.org/10.1089/tmj.2012.0011.

Soler, Jean-Karl, Inge Okkes, Maurice Wood, and Henk Lamberts (2008). "The Coming of Age of ICPC: Celebrating the 21st Birthday of the International Classification of Primary Care." *Family Practice* 25(4):312-317. Available at: https://doi.org/10.1093/fampra/cmn028.

Testern, Cecilie D., Lise Hestbæk, and Simon D. French (2015). The Use of Diagnostic Coding in Chiropractic Practice. *Chiropractic & Manual Therapies* 23:8. Available at: https://doi.org/10.1186/s12998-015-0051-1.

Wang, Yefeng, Jon Patrick, Graeme Miller, and Julie O'Hallaran (2008). A Computational Linguistics Motivated Mapping of ICPC-2 PLUS to SNOMED CT. *BMC Medical Informatics and Decision Making* 8 (Suppl 1 - October):S5. Available at: https://doi.org/10.1186/1472-6947-8-S1-S5.

WHO (2017). International Classification of Primary Care, Second Edition (ICPC-2). WHO. 2017. Available at: http://www.who.int/classifications/icd/adaptations/icpc2/en/.

World Organization of Family Doctors (Wonca), and Wonca International Classification Committee (WICC) (2017). *ICPC-2e* – English Version. Ehelse. No. 2.017. Available at: https://ehelse.no/icpc-2e-english-version.