Quality of Life through Quality of Information
J. Mantas et al. (Eds.)
IOS Press, 2012

doi:10.3233/978-1-61499-101-4-113

The Reproducibility of CLIF, a Method for Clinical Quality Indicator Formalisation

Kathrin DENTLER[a, c,1], Ronald CORNET[a], Annette TEN TEIJE[c], Kristien TYTGAT[b], Jean KLINKENBIJL[b] and Nicolette DE KEIZER[a]

[a] Dept. of Medical Informatics, [b] GIOCA,
Academic Medical Center, University of Amsterdam, The Netherlands
[c] Dept. of Computer Science, Vrije Universiteit Amsterdam, The Netherlands

Abstract. In order to be able to automatically calculate clinical quality indicators, we have proposed CLIF, a stepwise method for clinical quality indicator formalisation. Quality indicators are used for external accountability and hospital comparison. As clinical quality indicators are computed in a decentralised manner by the hospitals themselves, reproducibility of the formalisation method is essential to ensure the comparability of calculated values. Thus, we performed a case study to investigate the reproducibility of CLIF. Eight participants formalised the same sample quality indicator with the help of a web-based indicator-authoring tool that facilitates the application of CLIF. We analysed the results per step and concluded that the method itself leads to reproducible results. To further improve reproducibility, ambiguities in the indicator text must be clarified and trained experts are needed to encode clinical concepts and to specify the relations between concepts.

Keywords. Formalisation, Clinical Quality Indicators, Knowledge Representation, SNOMED CT

Introduction

A quality indicator is “a measurable element of practice performance for which there is evidence or consensus that it can be used to assess the quality, and hence change in the quality, of care provided” [1]. Calculated values are used internally to monitor and to improve the quality of delivered care, and externally to support patients and insurance companies in selecting hospitals of high performance. Ideally, clinical quality indicators are published in an unambiguous, standard representation, so that they can be computed automatically and are comparable among different institutions. We have presented CLIF, a stepwise method to formalise quality indicators into queries in [2]. In this paper, we report on a case study that we performed in order to investigate the reproducibility of CLIF. Our main research question was whether several persons who formalise the same quality indicator independently arrive at the same formalisation. We answered this question for each of CLIF’s steps. Any discrepancies were analysed to find the underlying cause.

[1] Corresponding Author: Kathrin Dentler. E-mail: k.dentler@amc.uva.nl

1. Methods

The case study is based on our previously proposed indicator formalisation method *CLIF* [2], which consists of eight steps. CLIF is applicable to process and outcome indicators expressed as proportions in general, but for testing its reproducibility, we focused on only one evidence-based process indicator defined by the Dutch healthcare inspectorate: “Number of examined lymph nodes after resection of a primary colonic carcinoma”. We chose this indicator because it is important in the domain of gastrointestinal oncology and because it is time-consuming to calculate manually as it requires data from several sources. When lymph nodes are examined after resection of a primary colonic carcinoma, at least 10 lymph nodes should be examined, and the indicator measures the proportion of patients for whom this is the case:

Numerator: Number of patients who had 10 or more lymph nodes examined after resection of a primary colonic carcinoma.
Denominator: Number of patients who had lymph nodes examined after resection of a primary colonic carcinoma.
Exclusion criteria: Previous radiotherapy and recurrent colonic carcinomas.
Reporting year: 2010

We created a *web-based indicator-authoring tool* to facilitate the formalisation process by leading users through the method step by step. The formalisation is performed against a problem-oriented information model with the central concepts “diagnosis” and “procedure”. The final result of the formalisation process is a query that is based on the information model. Our *test group* consisted of eight Master students in Medical Informatics. In an initial session, they were introduced to quality indicators, CLIF, the information model of our problem-oriented patient record and to SNOMED CT. They were trained on how to use the web-based tool and on how to search for SNOMED CT concepts in Snow Owl (http://www.b2international.com/portal/snow-owl)

Reference Standard. We developed a reference standard to measure the quality of the results of our participants. We studied the literature on which the indicator is based, consulted the institution that developed the indicator and organised a consensus meeting with medical informatics experts and clinical domain experts. **Table 1** shows the steps of CLIF and the developed reference standard.

Table 1. Steps of CLIF and the reference standard for the sample indicator.

Step	**CLIF**	**Reference standard for the sample indicator**
1)	Extract clinical concepts (e.g., diagnoses, procedures) from the indicator text. Search for matching concepts in a medical terminology using standard terminology browsing tools.	Table 2 shows the five relevant concepts from the indicator text with their correct encodings (emphasised). For example, the procedure “lymph nodes examined” from the indicator text is encoded by the SNOMED CT concept “Examination of lymph nodes”.
2)	The SNOMED CT concepts from *step 1* need to be related to the concepts of the information model. Finally, the relations between assigned concepts of the information model are defined.	All five SNOMED CT concepts encoded in step 1 have to be assigned to the correct concepts of the information model, i.e. SNOMED CT finding/disease concepts to the database table “diagnosis”, and procedure concepts to the database table “procedure”. To maintain a problem-oriented information model, all procedures should be related to diagnoses: lymph node examination, colectomy and radiotherapy have to be related to the diagnosis primary colonic carcinoma. The concept containing the number of examined lymph nodes should be related to the procedure lymph node examination.
3)	Temporal constraints are formalised.	The reporting year 2010 needs to be defined and related to “lymph node examination”, as this is the central procedure of the indicator. We expect two constraints to define that the lymph node examination has been after the start and before the end of the reporting year. We also expect two constraints that formalise the constructs “lymph

		nodes examined *after* resection" and "previous (i.e. *before* the colectomy) radiotherapy".
4)	Numeric constraints are formalised.	The only numeric constraint in the indicator is that the number of examined lymph nodes must be greater than or equal to 10.
5) & 6)	In *step 5*, Boolean constraints are formalised. In *step 6*, Boolean connectors can be used to group constraints.	There are no Boolean constraints in this indicator, and no constraints that have to be grouped by Boolean connectors.
7)	Exclusion criteria are defined.	Here, "radiotherapy", "recurrent colonic carcinoma" and the temporal constraint for "previous radiotherapy" have to be excluded.
8)	Constraints that only aim at the numerator are identified.	There is one constraint that only aims at the numerator: the numeric constraint that expresses that the number of examined lymph nodes should be higher than or equal to 10.

The reliability of agreement between the participants for encoding the concepts in SNOMED CT is measured as Fleiss' kappa and calculated in R.

2. Result

Figure 1 visualises the quality of the participants' solutions in terms of adherence to the reference standard per step.

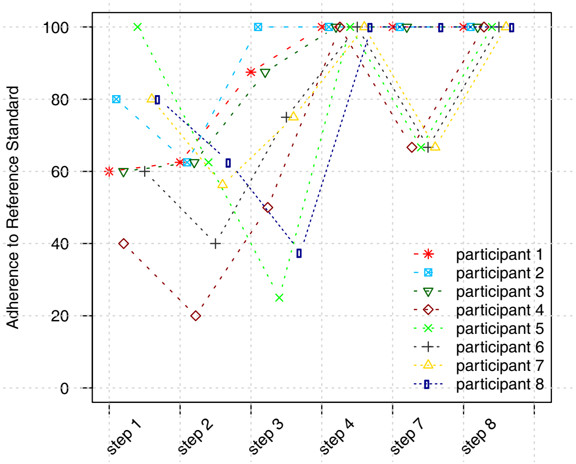

Figure 1. Quality of participants' solutions in terms of adherence to the reference standard. Each participant can reach up to 100 per cent for each step: To quantify the participants' solutions in *step 1*, each encoded concept that meets the reference standard receives 20%. In *step 2*, each correctly assigned concept receives 10%. Correct relations receive 12.5%, and solutions that use un-assigned concepts of the information model receive 6.25%. We do not penalise unnecessary relations. For *step 3*, all correct constraints receive 25% and the questionable ones 12.5%. Participants who formalised the numeric constraint in *step 4* reach 100%. Each constraint correctly excluded in *step 7* receives 33.33%, and participants who identified the constraint that only aims at the numerator in *step 8* receive 100%.

Step 1) All participants intended to encode exactly the five concepts contained in the reference standard. Seven of the eight participants entered five SNOMED CT concepts, and one entered four. The participants entered 9 different SNOMED CT concepts to encode these 39 concepts. All entered SNOMED CT concepts are subclasses of the two SNOMED CT concepts disease and procedure. **Table 2** gives an overview. The reliability of agreement between the participants for encoding these concepts in SNOMED CT is 0.754 ($p < 0.01$) according to Fleiss' kappa. This can be interpreted as substantial agreement.

Step 2) 36 out of 39 SNOMED CT concepts from *step 1* have been related to the correct concepts of the information model. Regarding the second substep, six of the eight participants related the colectomy to the primary colonic carcinoma. No participant has entered the three remaining relations contained in the reference standard.

Step 3) Six participants defined the reporting year. Five of them related it to the lymph node examination and one to an undefined procedure. Seven participants formalised the construct "lymph nodes examined after resection", while only four participants formalised the "previous radiotherapy". "Previous" was interpreted two times as having been carried out before the lymph node examination and one time before the colectomy. Another participant defined "previous radiotherapy" as having been performed before the start of the reporting year. This is questionable, as the radiotherapy might have taken place in the reporting year and before the colectomy.

Table 2. SNOMED CT concepts encoded by participants.

Indicator Text	**(Number of Participants) SNOMED CT Concept**	**Comment**
lymph nodes examined	(8) *Examination of lymph nodes*	Correct according to reference standard.
resection of a primary colonic carcinoma	(8) *Colectomy*	Correct according to reference standard.
radiotherapy	(7) *Radiation oncology AND/OR radiotherapy* (1) Radiation therapy procedure or service	Correct according to reference standard. Subconcept of correct concept. Contains only unreasonable subconcepts (e.g. "Disposal of radioactive source").
primary colonic carcinoma	(5) Carcinoma of colon (3) *Primary malignant neoplasm of colon*	Subconcept of correct concept. Defined via the associated morphology "Carcinoma, no subtype", which does not include specific carcinomas (e.g. adenocarcinoma) that should be included. Correct according to reference standard.
recurrent colonic carcinoma	(4) Secondary malignant neoplasm of colon (2) *Local recurrence of malignant tumor of colon* (1) Recurrent basal cell carcinoma	Sibling of correct concept. Synonym of metastasis in SNOMED CT; not related to recurrence. Correct according to reference standard. Skin carcinoma and thus not correct.

Step 4) Each of our eight participants defined the numeric constraint that we identified in the reference standard.

Step 7) All of the eight participants excluded the assigned concept "radiotherapy". Seven participants excluded "recurrent colonic carcinoma", and one excluded "carcinoma of colon". The participants also excluded all four temporal constraints that refer to "previous radiotherapy" and that have been formalised in *step 3*.

Step 8) All participants correctly identified the numeric constraint as only aiming at the numerator.

3. Discussion

We found that our eight participants could use CLIF reproducibly to formalise a sample quality indicator. For *step 1,* we concluded that detecting diagnoses and procedures in natural language text is a reproducible task. In contrast, encoding these concepts can lead to varying results. This task is complex due to the large size of medical terminologies. For example, SNOMED CT contains more than 311,000 hierarchically organised concepts, with many similar, interrelated concepts, making it hard to choose among them. Tools are required to support users in selecting the correct concepts. For *step 2*, we concluded that assigning the concepts to the information model is reproducible. However, our participants did not relate the assigned concepts of the information model as intended. This is due to insufficient knowledge of the employed information model. The reproducibility of *step 3* was lower than expected. This can be ascribed to the ambiguities of the indicator: it is not clear which events should occur in the reporting year and which event(s) the radiotherapy should precede. *Step 4* and *step*

8 were reproducible in our case study. In *step 7*, all participants who defined the constraint for "previous radiotherapy" also excluded it. In conclusion, CLIF itself leads to reproducible results, but the difficulty of encoding clinical concepts, defining relations between assigned concepts of the information model and ambiguities in the indicator text have a negative impact on its reproducibility.

Limitations. The main limitation of our study is that we only worked with one quality indicator, which did not require two (steps 5 and 6) out of CLIF's eight steps. Likewise, more participants would have been preferable. Finally, our results might have been biased by the choice of participants.

Related Work. Four clinical guidelines have been encoded into an early version of GLIF, the GuideLine Interchange Format, by two encoders each. The authors found that "different individuals produced different encodings as a result of different modeling choices, different representations of criteria given the use of narrative text in the current version of GLIF, and selection of different terminology for data elements in the absence of standards for clinical vocabulary and data models" [3]. We removed some of these obstacles in our case study: the authoring tool restricts possible formalisations, and we employed a standard terminology together with a common basic problem-oriented information model. Please note that later versions of GLIF adopt both standard terminologies and data models. An evaluation of the cognitive processes used in encoding guidelines in GLIF led to the conclusion that teams consisting of both clinicians and experts in computer-based representations produce better formalisations than individuals of either type working alone [5]. Medlock et al. [4] propose the 7-step Logical Elements Rule Method LERM to assess and formalise clinical rules, which are derived from quality indicators, for decision support. LERM has been validated empirically for inter-user reliability by comparing the results of two assessors who independently applied LERM on 16 rules. LERM was shown to be reliable provided that the users agree on a terminology and on when the rule will be evaluated.

Our main recommendations to increase the reproducibility of CLIF are: institutions that develop quality indicators should publish them together with sets of well-defined concepts from a standard terminology. Likewise, indicators have to be formulated as unambiguously and precisely as possible, so that they can be formalised and computed automatically. This is especially important with regard to temporal relations. The application of CLIF requires the cooperation of clinical domain experts to resolve ambiguities and medical informatics experts who are trained in clinical encoding and in the employed information model.

References

[1] Lawrence M, Olesen F. Indicators of Quality in Health Care. European Journal of General Practice. 1997;3(3):103–108

[2] Dentler K, ten Teije A, Cornet R, de Keizer N. Towards the Automated Calculation of Clinical Quality Indicators. Knowledge Representation for Health-Care LNCS 2012; 6924:51–64.

[3] Ohno-Machado L, Gennari J, Murphy SN, Jain NL, Tu SW, Oliver DE, Pattison-Gordon E, Greenes RA, Shortliffe EH, Barnett GO. The GuideLine Interchange Format: A Model for Representing Guidelines. JAMIA. 1998; 5(4):357–372.

[4] Medlock S, Opondo D, Eslami S, Askari M, Wierenga P, de Rooij SE, Abu-Hanna A. LERM (Logical Elements Rule Method): A method for assessing and formalizing clinical rules for decision support. International Journal of Medical Informatics. 2011;80(4):286–95.

[5] Patel VL, Allen VG, Arocha JF, Shortliffe EH. Representing Clinical Guidelines in GLIF: Individual and Collaborative Expertise. JAMIA. 1998;5(5):467–483.