Towards the Automated Calculation of Clinical Quality Indicators

Kathrin Dentler[1,2,*], Annette ten Teije[1], Ronald Cornet[2], and Nicolette de Keizer[2]

[1] Dept. of Computer Science, Vrije Universiteit Amsterdam, The Netherlands
[2] Dept. of Medical Informatics, Academic Medical Center, University of Amsterdam, The Netherlands
k.dentler@vu.nl

Abstract. To measure the quality of care in order to identify whether and how it can be improved is of increasing importance, and several organisations define quality indicators as tools for such measurement. The values of these quality indicators should ideally be calculated automatically based on data that is being collected during the care process. The central idea behind this paper is that quality indicators can be regarded as semantic queries that retrieve patients who fulfil certain constraints, and that indicators that are formalised as semantic queries can be calculated automatically by being run against patient data. We report our experiences in manually formalising exemplary quality indicators from natural language into SPARQL queries, and prove the concept by running the resulting queries against self-generated synthetic patient data. Both the queries and the patient data make use of SNOMED CT to represent relevant concepts. Our experimental results are promising: we ran eight queries against a dataset of 300,000 synthetically generated patients, and retrieved consistent results within acceptable time.

Keywords: Quality Indicators, Clinical Data, Formalisation of Clinical Quality Indicators, Semantic Web Reasoning, SPARQL, SNOMED CT.

1 Introduction

A quality indicator[1] 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" [9]. Quality indicators can be related to structure, process or outcome. According to Donabedian, structure denotes the attributes of the settings in which care occurs. Process denotes what is actually done in giving and receiving care, and outcome denotes the effects of care on the health status of patients and populations [5]. Process and outcome

* Corresponding author.

[1] The term quality indicator is used interchangeably with clinical / medical indicator / measure in this paper. However, as most measures are only indicators of quality, the term indicator is preferable [10].

D. Riaño, A. ten Teije, and S. Miksch (Eds.): KR4HC 2011, LNAI 6924, pp. 51–64, 2012.

indicators typically average over specific populations, and are often expressed by a fraction. The denominator consists of the relevant cohort of patients to whom the indicator applies, and the numerator of those patients contained in the denominator for which criteria that indicate (high or low) quality of care are fulfilled. Both for the population of the denominator and numerator, inclusion and exclusion criteria can apply.

Clinical quality indicators are typically being developed and released by governments, scientific associations, patient associations or insurance companies. They are calculated based on patient data within hospitals, and the obtained results are reported back to the indicator-releasing organisations. The increasing number of indicators makes their manual calculation difficult and time-consuming. Furthermore, indicators that are released in natural language need to be interpreted locally, which is error-prone due to the inherent ambiguity of natural language. Therefore, quality indicators should ideally be released in an unambiguous, machine-processable, formal representation in order to automatically calculate comparable values.

In this paper, we regard quality indicators as semantic queries against patient data, and propose a preliminary method for their formalisation into semantic queries. We prove the concept by applying exemplary formalised queries on self-generated coded data consisting of 300,000 patients. The next Section 2 presents our approach, and Section 3 our formalisation method. We detail the generation of synthetic patient data in Section 4, and present our experimental results in Section 5. We end the paper by discussing related work in Section 6, future work in Section 7 and our conclusions in Section 8.

2 Approach

Our test set of quality indicators (see appendix) contains four indicators that have been released in natural language and stem from the domain of gastrointestinal cancer surgery, but in principle, we aim for a domain-independent approach. We investigate the feasibility of formalising the set of indicators into SPARQL queries[2]. The exemplary SPARQL query below retrieves all instances of type patient (the SNOMED CT code for "patient" is `SCT_116154003`). The SELECT clause defines the only variable that is to be retrieved as result (i.e. ?patient), and the WHERE clause defines a triple pattern which contains the same variable and is to be matched against the data graph.

```
SELECT ?patient
WHERE {
  ?patient a sct:SCT_116154003 .
}
```

Our proposed formalisation method consists of 8 steps: 1) to encode relevant concepts from the indicator by concepts from a terminology, 2) to define the information model, and 3) to 5) to formalise temporal, numeric and boolean constraints as SPARQL FILTERs. Step 6) is to group constraints by boolean

[2] http://www.w3.org/TR/sparql11-query/

connectors, step 7) to identify exclusion criteria and step 8) to identify constraints that only aim at the numerator, in order to construct the denominator by removing these constraints. All steps are explained in Section 3.

To test the formalised queries, we synthetically generated patient data that is represented in OWL 2[3], allowing for automated reasoning and semantic interoperability. We employ SNOMED CT [3] concepts from the July 2010 version to describe both the query variables (step 1 of our method) and our patient data. Typically, patient data is very detailed, but quality indicators query for groups of patients on a less granular level. We employ Semantic Web reasoning to bridge this gap by inferring subclass relationships. For example, generated rectum cancer patients are undergoing the procedures "Stapled transanal resection of rectum" or "Wedge resection of rectum", which are both subclasses of "Resection of rectum". To calculate an indicator, we query for all patients with a procedure of type "Resection of rectum" and retrieve all patients with subclasses of this procedure by automated reasoning.

3 Formalisation of Quality Indicators

This section describes our formalisation method. As the numerator is always a subset of the denominator, and is thus restricted by more constraints, we first formalise the numerator and afterwards construct the denominator from it by removing constraints. We formalised a set of four quality indicators (see appendix, referred to as I1 - I4). In the following, we present our method by formalising the exemplary process indicator "Number of examined lymph nodes after resection" (I1). The clinical background of the indicator is a colon cancer guideline that states: "A minimum of 10 lymph nodes is recommended to assess a negative lymph node status". The original version of the indicator is:

> **I1: Number of examined lymph nodes after resection (process indicator)**
> *Numerator: number of patients who had 10 or more lymph nodes examined after resection of a primary colon carcinoma.*
>
> *Denominator: number of patients who had lymph nodes examined after resection of a primary colon carcinoma.*
>
> Exclusion criteria: Previous radiotherapy and recurrent colon carcinomas

Step 1: Encoding of relevant concepts from the indicator by concepts from a terminology. The first step of our method is to extract all required concepts from the indicator, and to find the corresponding concepts in a terminology, in our case SNOMED CT. We perform this step first because the concepts are the building blocks for further formalisation. In SPARQL, we encode the query variables based on those concepts:

```
?patient a sct:SCT_116154003 .
```

[3] http://www.w3.org/TR/owl2-overview/

Step 2: Definition of the information model. Subsequently, we define the information model, i.e. how the resources are related to each other. This step could be automated once a standard information model is employed. In SPARQL:

```
?patient ehrschema:hasDisease ?coloncancer .
```

Step 3: Formalisation of temporal constraints (FILTER). The next step is to formalise temporal constraints. This step helps us to discover an ambiguity: the indicator does not state explicitly what should be included the reporting year. It could be for example the resection of the carcinoma or the lymph node examination. Because the indicator aims at the number of examined lymph nodes, we assume the latter. One of the temporal relationship between two query variables in this indicator states that the lymph node examination has to follow the colectomy. These constraints are expressed as FILTERs in SPARQL. FILTERs restrict solutions to those for which the filter expressions evaluate to true:

```
FILTER ( ?lymphnodeexaminationdate > "2010-01-01T00:00:00+02:00"^^xsd:dateTime )
FILTER ( ?lymphnodeexaminationdate < "2011-01-01T00:00:00+02:00"^^xsd:dateTime )
FILTER ( ?lymphnodeexaminationdate > ?colectomydate)
```

Step 4: Formalisation of numeric constraints (FILTER). The only numeric constraint contained in the indicator is that the number of examined lymph nodes has to be 10 or more. In SPARQL:

```
FILTER ( ?numberexaminedlymphnodes >= 10 )
```

Step 5: Formalisation of boolean constraints (FILTER). The exemplary indicator does not contain boolean constraints. However, the indicator "Participation in Dutch Surgical Colorectal Audit" (DSCA, I2) asks for patients for which data has been delivered to the DSCA. In SPARQL:

```
FILTER ( ?dataDeliveredToDSCA = true)
```

Step 6: Grouping of constraints by boolean connectors. All elements of the constructed SPARQL query are connected by logical conjunctions. However, some queries require logical disjunctions. An example is again I2, which asks for surgical resections of a colorectal carcinoma situated in colon *or* rectum:

```
{ ?cancer  a sct:coloncancer . ?operation a sct:colectomy }
UNION
{ ?cancer a sct:rectumcancer . ?operation  a sct:resectionrectum }
```

Step 7: Identification of exclusion criteria (FILTER). One of the exclusion criteria of the example indicator is "previous radiotherapy". Thus, we exclude all patients who underwent radiotherapy before the lymph node examination. All criteria that are not explicitly identified as exclusion criteria are inclusion criteria.

```
FILTER NOT EXISTS {
  ?radiotherapy a sct:SCT_108290001 .
  ?patient ehrschema:hasProcedure ?radiotherapy .
  ?radiotherapy ehrschema:procedureDate ?radiotherapydate .
  FILTER ( ?lymphnodeexaminationdate > ?radiotherapydate)
}
```

Step 8: Identification of constraints that only aim at the numerator. In this step, the numerator is already formalised, and constraints are removed to construct the query for the denominator. In order to do so, it is important to be aware of the clinical intent of the indicator. Regarding the example indicator, it is considered good practice to examine 10 or more lymph nodes. Therefore, the only constraint that is removed to construct the denominator is: "number of examined lymph nodes >= 10".

Resulting SPARQL query (Numerator)

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ehrschema: <http://apdg.net/owl/schema/>
PREFIX sct: <http://www.ihtsdo.org/>

SELECT ?patient
WHERE {

# step 1)
?patient a sct:SCT_116154003 .
?coloncancer a sct:SCT_93761005 .
?colectomy a sct:SCT_23968004 .
?lymphnodeexamination a sct:SCT_284427004 .

# step 2)
?colectomy sct:SCT_47429007 ?coloncancer . # SCT_47429007 = associated with
?patient ehrschema:hasDisease ?coloncancer .
?patient ehrschema:hasProcedure ?colectomy .
?colectomy ehrschema:procedureDate ?colectomydate .
?patient ehrschema:hasProcedure ?lymphnodeexamination .
?lymphnodeexamination ehrschema:procedureDate ?lymphnodeexaminationdate .
?lymphnodeexamination ehrschema:hasNumber ?numberexaminedlymphnodes .

# step 3)
FILTER ( ?lymphnodeexaminationdate > "2010-01-01T00:00:00+02:00"^^xsd:dateTime )
FILTER ( ?lymphnodeexaminationdate < "2011-01-01T00:00:00+02:00"^^xsd:dateTime )
FILTER ( ?lymphnodeexaminationdate > ?colectomydate)

# step 4); needs to be removed to construct the denominator (step 8)
FILTER ( ?numberexaminedlymphnodes >= 10 )

# step 7)
FILTER NOT EXISTS {
  ?radiotherapy a sct:SCT_108290001 .
  ?patient ehrschema:hasProcedure ?radiotherapy .
  ?radiotherapy ehrschema:procedureDate ?radiotherapydate .
  FILTER ( ?lymphnodeexaminationdate > ?radiotherapydate)
}}
```

Regarding the order of the steps, step 1) and 2) should be carried out first, because they formalise the building blocks that are used in subsequent steps. Steps 6) - 8) should be carried out last, because they build on previously defined constraints. Steps 3) to 5) can be performed in the preferred order of the user.

Experiences during formalisation. We succeeded in formalising all four quality indicators included in our example set as SPARQL queries with the method as described above, and the formalisation process was relatively straightforward. The only construct that is not directly expressible in SPARQL is: "number of re-interventions during the same admission or during 30 days after the resection (choose longest interval)" (I4), because there is no function to subtract dates from each other in SPARQL. This is clearly an insufficiency. Two possible options to circumvent this problem are to implement a custom extension function or to first query for all patients who had a re-intervention and then to apply the filter on the retrieved results. Both solutions need to be implemented locally (extension functions have to be implemented for the triple store that is being queried, and results need to be filtered where the data is retrieved), and thus allow for the introduction of implementation errors and limit interoperability.

We found a high coverage of SNOMED CT with respect to the colorectal cancer surgery domain. The only concept that we could not encode was the exclusion criterion "Transanal Endoscopic Microsurgery (TEM)" (I3 and I4). We excluded "Stapled transanal resection of rectum", "Transanal disk excision of rectum" and "Transanal resection of rectum and anastomosis using staples" instead. None of these replacements are explicitly "endoscopic". Alternatives would have been to post-coordinate the concept or to employ a concept from another terminology.

We did not implement subtleties such as the presence of a radiologist, a radiotherapist, a surgeon, an oncologist, a colon, stomach and liver physician and a pathologist in a multidisciplinary meeting (I3). This would in principle be possible, but we argue that it is unrealistic to expect that meeting protocols document the presence of individual persons. Another concept that we did not implement is the definition of re-intervention. We employed the SNOMED CT concept "Reoperation" instead, and defined that it must be associated to the same carcinoma that the first operation was associated to.

We noticed a considerable variability in the natural language descriptions of the indicators contained in our test set. For example, all carcinomas should be primary and not recurrent. This is expressed in four different ways for four different indicators: I1) resection of a primary colon carcinoma (numerator and denominator); Exclusion criterion: recurrent colon carcinomas, I2) only count primary carcinomas (numerator and denominator), I3) Exclusion criterion: recurrent rectum carcinomas, I4) Inclusion criterion: Primary colorectal carcinoma = first presentation of a colorectal carcinoma (thus not recurrent); might be the second or next primary presentation.

We encountered several ambiguities and conclude that the expertise of a domain expert is indispensable during the formalisation process.

Another observation is that many concepts occur in several indicators (e.g. colectomy), but there are also concepts that only occur in one indicator (e.g. lymph node examination). Table 1 shows the concepts and data items required to calculate the numerators (and thus also the denominators) contained in our quality indicator set. Similarly to the concepts, some filter patterns occur in all indicators, and others are indicator-specific. Table 2 gives an overview of the numbers of constraints that are required to calculate the numerators of the indicators. We conclude that many patterns can be re-used once they are created.

Table 1. Concepts required to calculate quality indicators

Concept	**I1 (lymph nodes)**	**I2 (DSCA)**	**I3 (meeting)**	**I4 (reoperation)**
patient (SCT_116154003)	x	x	x	x
associated with (SCT_47429007)	x	x	x	x
lymph node exam. (SCT_284427004)	x			
lymph node examination date (date)	x			
number of examined lymph nodes (int)	x			
radiotherapy (SCT_108290001)	x			
radiotherapy date (date)	x			
pr. colon cancer (SCT_93761005)	x	x		x
pr. rectum cancer (SCT_93984006)		x	x	x
colectomy (SCT_23968004)	x	x		x
colectomy date (date)	x	x		x
resection rectum (SCT_87677003)		x	x (plus subconcepts)	x (plus subconcepts)
resection rectum date (date)		x	x	x
delivered to DSCA (boolean)		x		
multidisc. meeting (SCT_312384001)			x	
multidisc. meeting date (date)			x	
re-operation (SCT_261554009)				x
re-operation date (date)				x
polypectomy (SCT_82035006)				x
discharge date (date)				x

Table 2. Numbers of SPARQL filters required to calculate quality indicators

Filter	**I1 (lymph nodes)**	**I2 (DSCA)**	**I3 (meeting)**	**I4 (reoperation)**
Temporal Constraints (step 3)	4 (operation within reporting year; examination after colectomy; previous radiotherapy)	2 (operation within reporting year)	3 (operation within reporting year; meeting before resection)	5 (operation and reoperation within reporting year; operation before reoperation)
Numeric Constraints (step 4)	1 (number lymph nodes examined)	-	-	-
Boolean Constraints (step 5)	-	1 (data delivered to DSCA)	-	-
Exclusion Criteria (step 6)	1 (no previous radiotherapy)	-	3 (excluded TEM concepts)	4 (excluded TEM concepts and polypectomy)

4 Generation of Data for all Indicators

We generated synthetic patient data in order to be able to test our formalised queries. It consists of an OWL schema that describes the data needed to calculate the exemplary indicators (TBox, i.e. terminological background knowledge), and the patient data (ABox, i.e. knowledge about individuals). We generated both the OWL schema and the patient data in OWL 2 with the OWL API [6]. Figure 1 shows the OWL schema. We deliberately kept this model as simple as possible (it consists of 25 axioms), and it reflects the information model as employed by the SPARQL queries. The OWL classes "Patient", "Procedure", "Disease" and "Examination of lymph nodes" are SNOMED CT concepts. In the schema, the classes are represented by their SNOMED CT identifiers, e.g. `sct:SCT_116154003` for "Patient". We also added the SNOMED CT concepts "Primary malignant neoplasm of colon" , "Secondary malignant neoplasm of colon", "Primary malignant neoplasm of rectum" and "Secondary malignant neoplasm of rectum", which are all Diseases, and the Procedures "Colectomy", "Resection of rectum", "Radiation oncology AND/OR radiotherapy", "Multi-disciplinary assessment" and "Reoperation".

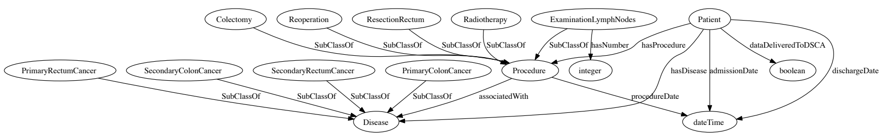

Fig. 1. OWL Schema

The data generator generates an arbitrary number of patients as instances of the OWL Class "Patient". All generated patients are colon cancer (50 percent) or rectum cancer (50 percent) patients who underwent colectomy or resection of rectum during a random operation date within the years 2009 to 2011 (we assume that the reporting year is 2010). The malignant neoplasm is primary in 50 percent of the cases, otherwise it is secondary. All generated rectum cancer patients receive a random subclass of the SNOMED CT concept "Resection of rectum" as procedure. The data generator retrieves those subclasses with the help of FaCT++ [15]. Examples are "Stapled transanal resection of rectum" or "Wedge resection of rectum". Patients are admitted to the hospital one day before the operation and discharged between 1 and 60 days after the operation. 10 percent of the patients are re-operated between 1 and 60 days after the first operation. A patient has a lymph node examination with a probability of 50 percent at a random date within 60 days after the operation, with a random number (between 1 and 20) of examined lymph nodes. With a probability of 20 percent, the patient received radiotherapy at a random date within 60 days before the operation. Rectum cancer patients are discussed in a multidisciplinary meeting at a random date within 60 days before the operation with a probability of 80 percent and for all patients, data is sent to the DSCA with a probability of 90 percent. The defined temporal constraints result in radiotherapy always taking place before a lymph node examination, and a multidisciplinary meeting always before the operation. All probabilities are chosen arbitrarily.

Figure 2 shows an exemplary generated patient, and Figure 3 an extract of the same patient in OWL Functional Syntax. The data generator produces around 15 triples per patient, thus our ABox for 300,000 patients consists of over 4 million triples (4,530,578).

5 Experimental Results

In this section, we present our experimental results with respect to the calculation of the formalised indicators, i.e. the execution of the SPARQL queries

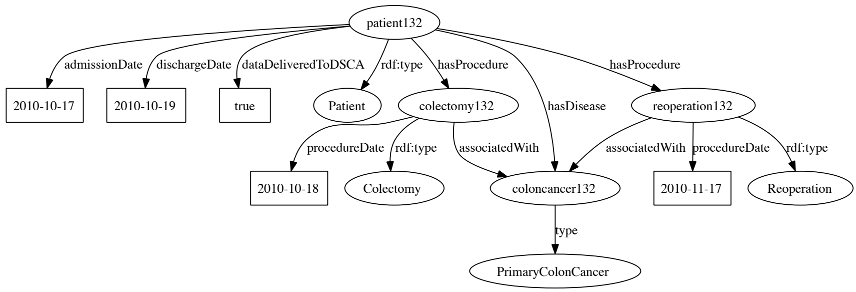

Fig. 2. Synthetically Generated Patient

```
Declaration(NamedIndividual(data:patient132))
ClassAssertion(sct:SCT_116154003 data:patient132)
ObjectPropertyAssertion(ehrschema:hasDisease data:patient132 data:coloncancer132)
ObjectPropertyAssertion(ehrschema:hasProcedure data:patient132 data:reoperation132)
ObjectPropertyAssertion(ehrschema:hasProcedure data:patient132 data:colectomy132)
DataPropertyAssertion(ehrschema:admissionDate data:patient132 "2010-10-17T05:49:20+02:00"^^xsd:dateTime)
DataPropertyAssertion(ehrschema:dataDeliveredToDSCA data:patient132 "true"^^xsd:boolean)
DataPropertyAssertion(ehrschema:dischargeDate data:patient132 "2010-10-19T05:49:20+02:00"^^xsd:dateTime)
```

Fig. 3. Synthetically Generated Patient in OWL Functional Syntax

against the generated patient data. We derived the closure of SNOMED CT with CB [7], the fastest reasoner currently available for this nomenclature [4]. Then, we loaded the closure, our OWL schema and the patient data into BigOWLIM 3.5 [8], which is optimised for fast SPARQL evaluation and was allowed a maximum of 6GB memory. We employed openRDF Sesame 2.4 [2], which supports SPARQL 1.1[4] query features such as expressions, aggregates and negation.

We ran two queries per indicator: one for the numerator and one for the denominator. For the construct "number of re-interventions during the same admission or during 30 days after the resection (choose longest interval)" (I4), we chose to filter the final results from the results returned by the query and measured the runtime including this filtering. Table 3 shows the number of retrieved patients for the numerators and denominators of our queries, and the calculated percentage for each indicator. The last two rows of the table contain the runtimes for the queries, averaged over 100 runs. All queries are processed within seconds. As the calculation of quality indicators is not time-critical, the runtimes are acceptable.

Table 3. Number of results and runtimes in seconds

Data Item	**I1(lymph nodes)**	**I2(DSCA)**	**I3(meeting)**	**I4(reoperation)**
numerator	5,449	44,878	17,439	2,713
denominator	9,898	49,848	21,807	49,848
percent	55%	90%	80%	0.5%
runtime numerator	14.28	25.12	17.74	9.88
runtime denominator	15.90	25.71	15.43	41.36

[4] `http://www.w3.org/TR/sparql11-query/`

We checked whether the experimental results are correct by comparing them to the results that we expected based on the probabilities that were used for data generation. For example, the DSCA indicator applies to primary colon and rectum cancer patients, i.e. 50% of our population (150,000). One third of these patients (50,000) is expected to have been operated in 2010, and 90% of the data is sent to the DSCA (45,000). The corresponding query retrieved 44,878 patients, which is comparable. Also the percentages are consistent: for example, the data generator produced a random number between 1 and 20 examined lymph nodes, and 55% of the examinations inspected 10 or more lymph nodes. The fact that we obtained consistent results within acceptable time based on the formalised SPARQL queries and synthetically generated patient data proves the concept and shows that the queries are well-formalised.

6 Related Work

6.1 Formalisation of Quality Indicators

In the following, we discuss a method to formalise goals [14], and a formalisation method for clinical rules [11]. As they do not consider numerators and denominators and in- and exclusion criteria, which are the core elements of quality indicators, neither of the methods is directly applicable to our use case. Thus, we follow our own approach (Section 3) that re-uses steps of these methods wherever applicable. Both methods are gradual, and we believe that this is essential in order to preserve the clinical intent of indicators during their formalisation.

Stegers et al. [14] propose a 5-step method to translate goals (e.g. quality indicators) from natural language to the formalism of a verification tool. A domain expert is involved to guarantee the correctness of the result. The authors contribute a conceptual goal model, which serves as a common frame of reference for all involved experts and can be expressed in a formal language. Their method consists of the following steps: 1) Reduction: explicitly describe the clinical intent of the indicator. 2) Normalisation: rewrite the goal in terms of the goal model. This disambiguates temporal constraints. 3) Formalisation: transform the structured natural language version to a formalised version in GDL (Goal Definition Language). 4) Attachment: formalise the natural language parts with concepts available in the process model. 5) Translation: transform GDL to the logic of the verification tool. This step should be strictly mechanical.

Elements of the method that we re-use are "Reduction" to make the clinical intent of the indicator explicit, which is needed to construct the denominator from the numerator in step 8) of our method, "Normalisation" in order to disambiguate temporal constraints in step 3) of our method and "Attachment", to encode relevant concepts and define the information model in step 1) and 2) of our method. "Formalisation" and "Translation" are not applicable.

Medlock et al. [11] propose the Logical Elements Rule Method (LERM), a 7-step method to transform clinical rules for use in decision support: (1) restate the rule proactively; (2) restate the rule as a logical statement (preserving key phrases); (3) assess for conflict between rules; (4) identify concepts which are not needed; (5) classify concepts as crisp or fuzzy, find crisp definitions corresponding to fuzzy concepts, and extract data elements from crisp concepts; (6) identify rules which are related by sharing patients, actions, etc.; (7) determine availability of data in local systems.

We re-use step (1) "restate the rule proactively" to make the clinical intent of the indicator explicit in step 8) of our method and step (5) "classify concepts as crisp or fuzzy, ..." to encode concepts, although we do not differentiate between crisp and fuzzy concepts, in step 1) of our method. Steps (3) "assess for conflict between rules" and (6) "identify rules which are related by sharing patients, actions, etc." relate several indicators. Because indicators are typically calculated independently from each other, these steps are not needed for our application scenario. Step (2) "restate the rule as a logical statement" is similar to step 6) of our method, which groups constraints by boolean connectors. Additionally, exclusion criteria are negated, and the elements of our SPARQL query are connected by logical conjunctions. Our method does not contain a step (4) "identify concepts which are not needed", as non-needed concepts do not need to be encoded. We consider step (7) "determine availability of data in local systems" to be part of the calculation of an indicator.

6.2 Calculation of Quality Indicators

Once an indicator has been formalised, it can be calculated based on patient data. Previous attempts to automatically calculate quality indicators include [17] and [12]. The main conclusion of [17] is that for automated chart reviews, more fully-structured and coded data would have to be entered by physicians. As we generate synthetic patients, we do not encounter this problem. The authors of [12] present a rule-based Analytics Engine that is capable of interpreting documents in the Health Quality Measures Format (HQMF)[5] and generating reports. HQMF is a machine-processable standard for representing health quality measures as electronic documents (eMeasures).

6.3 Indicators and Eligibility Criteria

In- and exclusion criteria are referred to as eligibility criteria [16] and are commonly employed not only for quality indicators, but also for protocols, guidelines, and clinical studies and trials. In the following, we describe two methods for clinical trial recruitment [1], [13] that are based on Semantic Web technologies. Similar to our approach, both methods employ a terminology. In contrast to our approach, they rely on SWRL or description logic queries instead of SPARQL. Besana et al. [1] showed that the automatic recruitment of patients who meet eligibility criteria of clinical trials is possible based on OWL and SWRL, the Semantic Web Rule Language[6]. They use the NCI ontology to represent both

[5] http://www.hl7.org/v3ballot/html/domains/uvqm/uvqm.html

[6] http://www.w3.org/Submission/SWRL/

patient data and the eligibility criteria. Patel et al. [13] demonstrated that clinical trial criteria can be formulated as description logic queries, which a reasoner can use together with SNOMED CT to infer implicit information that results in retrieving eligible patients.

7 Future Work

As we worked with arbitrary probabilities, the data produced by our data generator is not representative. With the help of a domain expert, it might have been possible to generate more meaningful clinical data. Furthermore, the use of self-generated data leads to avoiding common problems such as insufficient data quality and missing as well as irrelevant data items, but with respect to the difficulty of obtaining (large amounts of) real patient data we consider it to be useful to calculate first indicators as a proof of concept. In the future, we will work with real patient data that stems from several sources.

Our set of four exemplary quality indicators is not representative either. We will work with a larger, more diverse set of indicators in the future in order to further investigate the generalisability of our method. Another open question is whether quality indicators released in natural language are precise enough to be formalised. We will cooperate with domain experts in order to answer this question and to ensure that the clinical intent of the quality indicator is preserved during its formalisation.

8 Conclusions

We presented a 8-step method that is inspired by previously proposed methods [14], [11] to formalise quality indicators as SPARQL queries. The steps are: 1) to encode relevant concepts from the indicator by concepts from a terminology, 2) to define the information model, and 3) to 5) to formalise temporal, numeric and boolean constraints as SPARQL FILTERs. Step 6) is to group constraints by boolean connectors, step 7) to identify exclusion criteria and step 8) to identify constraints that only aim at the numerator, in order to construct the denominator by removing these constraints. Applying this method, we succeeded in formalising a set of four quality indicators into SPARQL queries.

We encountered one construct that is not directly expressible in SPARQL. Although this limits interoperability, the problem can be circumvented. We found a high coverage of SNOMED CT with respect to the colorectal cancer domain. We noticed variability and ambiguity in the original descriptions of the quality indicators and conclude that a domain expert is indispensable to ensure the clinical correctness of the formalised indicators. Finally, we observed that many concepts and filter patterns can be reused once they are formalised.

We proved the concept by running the SPARQL queries that resulted from the formalisation process against self-generated data that consisted of 300,000 synthetically generated patients, and retrieved results that are consistent with the generated data in acceptable time. We conclude that semantic queries are are a promising step towards the automated calculation of clinical quality indicators.

References

1. Besana, P., Cuggia, M., Zekri, O., Bourde, A., Burgun, A.: Using Semantic Web technologies for Clinical Trial Recruitment. In: Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J.Z., Horrocks, I., Glimm, B. (eds.) ISWC 2010, Part II. LNCS, vol. 6497, pp. 34–49. Springer, Heidelberg (2010)
2. Broekstra, J., Kampman, A., Van Harmelen, F.: Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema. In: Horrocks, I., Hendler, J. (eds.) ISWC 2002. LNCS, vol. 2342, pp. 54–68. Springer, Heidelberg (2002)
3. Cornet, R., de Keizer, N.: Forty years of SNOMED: a literature review. BMC Medical Informatics and Decision Making 8(suppl 1), S2 (2008)
4. Dentler, K., Cornet, R., ten Teije, A., de Keizer, N.: Comparison of reasoners for large ontologies in the OWL 2 EL profile. Semantic Web 2, 71–87 (2011)
5. Donabedian, A.: The Quality of Care: How Can It Be Assessed? JAMA (1988)
6. Horridge, M., Bechhofer, S.: The OWL API: A Java API for OWL ontologies. Semantic Web Journal (to appear), `http://www.semantic-web-journal.net/`
7. Kazakov, Y.: Consequence-driven reasoning for horn SHIQ ontologies. In: Proceedings of the 21st International Workshop on Description Logics, pp. 2040–2045 (2009)
8. Kiryakov, A., Ognyanov, D., Manov, D.: OWLIM – A Pragmatic Semantic Repository for OWL. In: Dean, M., Guo, Y., Jun, W., Kaschek, R., Krishnaswamy, S., Pan, Z., Sheng, Q.Z. (eds.) WISE 2005 Workshops. LNCS, vol. 3807, pp. 182–192. Springer, Heidelberg (2005)
9. Lawrence, M., Olesen, F.: Indicators of Quality in Health Care. European Journal of General Practice 3(3), 103–108 (1997)
10. Lilford, R., Mohammed, M.A., Spiegelhalter, D., Thomson, R.: Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. Lancet 363(9415), 1147–1154 (2004)
11. Medlock, S., Opondo, D., Eslami, S., Askari, M., Wierenga, P., de Rooij, S.E., Abu-Hanna, A.: LERM (Logical Elements Rule Method): A method for assessing and formalizing clinical rules for decision support. International Journal of Medical Informatics 80(4), 286–295 (2011)
12. Palchuk, M.B., Bogdanova, A.A., Jatkar, T., Liu, J., Karmiy, N., Housman, D., Einbinder, J.S.: Automating Quality Reporting with Health Quality Measures Format "eMeasures" and an Analytics Engine. In: AMIA Symposium Proceedings, page 1205 (2010)
13. Patel, C., Cimino, J., Dolby, J., Fokoue, A., Kalyanpur, A., Kershenbaum, A., Ma, L., Schonberg, E., Srinivas, K.: Matching Patient Records to Clinical Trials Using Ontologies. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) ASWC 2007 and ISWC 2007. LNCS, vol. 4825, pp. 816–829. Springer, Heidelberg (2007)
14. Stegers, R., ten Teije, A., van Harmelen, F.: From Natural Language to Formal Proof Goal. In: Staab, S., Svátek, V. (eds.) EKAW 2006. LNCS (LNAI), vol. 4248, pp. 51–58. Springer, Heidelberg (2006)
15. Tsarkov, D., Horrocks, I.: FaCT ++ Description Logic Reasoner: System Description
16. Weng, C., Tu, S.W., Sim, I., Richesson, R.: Formal representation of eligibility criteria: A literature review. Journal of Biomedical Informatics 43(3), 451–467 (2010)
17. Williams, C.A., Mosley-Williams, A.D., Overhage, J.M.: Arthritis Quality Indicators for the Veterans Administration: Implications for Electronic Data Collection, Storage Format, Quality Assessment, and Clinical Decision Support. In: AMIA Symposium Proceedings, pp. 806–810 (January 2007)

Appendix: Set of Quality Indicators

The indicators are released by the Dutch healthcare inspectorate and contained in the indicator set for 2011.

I1: Number of examined lymph nodes after resection (process indicator)
Numerator: number of patients who had 10 or more lymph nodes examined after resection of a primary colon carcinoma.

Denominator: number of patients who had lymph nodes examined after resection of a primary colon carcinoma.

Exclusion criteria: Previous radiotherapy and recurrent colon carcinomas

I2: Participation in Dutch Surgical Colorectal Audit (DSCA) (process indicator)
Numerator: number of surgical resections of a colorectal carcinoma situated in colon or rectum (only count primary carcinomas) for which data has been submitted to the Dutch Surgical Colorectal Audit.

Denominator: total number of surgical resections of a colorectal carcinoma situated in colon or rectum (only count primary carcinomas).

I3: Patients with rectum carcinoma who have been discussed in a preoperative multidisciplinary meeting (process indicator)
Numerator: Number of patients with rectum carcinoma who have been discussed in a preoperative multidisciplinary meeting.

Denominator: Number of patients with rectum carcinoma operated in the reporting year.

Inclusion criterion: Patients who have been operated in the reporting year due to a rectum carcinoma.

Exclusion criteria: Transanal Endoscopic Microsurgery (TEM) resections and recurrent rectum carcinomas.

The Dutch Surgical Colorectal Audit states that the presence of a radiologist, a radiotherapist, a surgeon, an oncologist, a colon, stomach and liver physician and a pathologist are required for a preoperative multidisciplinary meeting.

I4: Unplanned re-interventions after resection of a primary colorectal carcinoma (outcome indicator)
Numerator: number of re-interventions during the same admission or during 30 days after the resection (choose longest interval) in the reporting year.

Denominator: total number of primary resections of a colorectal carcinoma during the reporting year.

Inclusion criteria: Primary colorectal carcinoma = first presentation of a colorectal carcinoma (thus not recurrent); might be the second or next primary presentation.

Exclusion criteria: Transanal Endoscopic Microsurgery (TEM); Endoscopic and open polypectomy

This indicator comes with a list of definitions: Resection: surgical removal of colon segment where the colorectal carcinoma is situated. Re-intervention: re-operation in the abdomen or an intervention (possibly radiological) during which a complication in the abdomen is being treated (inclusive percutaneous incision and drainage, drainage via rectum, embolisations of bleedings in the abdomen, etcetera). Admission: the time which the patient spends in a hospital directly after the operation (the same hospital or another one where the patient has been referred to); can be longer than 30 days.