Development of an Ontological Model of Evidence for TRANSFoRm Utilizing Transition Project Data

Derek Corrigan[1], Jean-Karl Soler[2], Brendan Delaney[3]

[1]Royal College of Surgeons in Ireland, Dublin, Ireland
derekcorrigan@rcsi.ie
[2]Mediterranean Institute for Primary Care, Attard, Malta
jksoler@synapse.net.mt
[3]Kings College London, London, United Kingdom
brendan.delaney@kcl.ac.uk

Abstract. The development of decision support tools that assist clinicians effectively practice evidence-based-medicine in primary care is dependent on the development of formal models of clinical knowledge. These formal models are a pre-requisite for bridging the knowledge gap that exists between generation of research knowledge and its application in clinical practice. The TRANSFoRm project has developed formal ontological models to represent diagnostic clinical knowledege providing a basis for future development of diagnostic decision support tools. The conceptual validity of the developed models has been tested through representation of diagnostic clinical evidence obtained from literature sources and International Classification of Primary Care Second Edition (ICPC2) coded clinical evidence captured as part of the Transition project. The models provide a basis for future development of decision support tools as part of the on-going TRANSFoRm project. These tools can assist clinicians to formulate and quantify potential diagnoses based on diagnostic cues extracted from patient electronic health records.

Keywords: Ontology, Semantic Web, Evidence-Based-Medicine, Electronic Health Record, Decision Support

1 Introduction

The application of systematic and rigorous approaches to diagnosis through access to the latest available clinical research has long been advocated as one way of contributing to improving patient safety in family practice. The term 'evidence based medicine' has been widely associated with such approaches [1]. The effective practice of evidence based medicine implies the existence and use of an up-to-date repository of clinical knowledge. This can be used for interpretation of the diagnostic cues associated with a presenting patient (whether or not this evidence be in electronic format or written) [2]. The challenges in keeping a repository of diagnostic information up to

date are similar to the problems of keeping evidence of the effectiveness of treatments up to date. This manifests itself in a delay between the generation of new clinical knowledge from research activities and the timely dissemination of this knowledge into actual clinical practice [3]. Translational medicine advocates the quicker dissemination of research knowledge to clinical practice. It studies the pathways and mechanisms that may optimally provide for the translation of research knowledge into actionable knowledge in clinical practice [4]. One core area highlighted in the study of translational medicine and evidence based medicine has been the need for the development of more formal shared models and coding of clinical data [5]. This can enable quicker dissemination of actionable knowledge via electronic medical record systems. This paper describes how the TRANSFoRm project ('Translational Medicine and Patient Safety in Europe') is working to develop such formal models provided through a dynamically updateable ontology of coded clinical evidence that will support deployment as part of a broader translational medicine platform.

2 The TRANSFoRm Project

The TRANSFoRm project is a five year EU FP7 funded project involving the cooperation of over 20 academic and industry based European research partners. The aim of TRANSFoRm is to develop and evaluate an electronic infrastructure for the 'learning healthcare system' to support both research (epidemiology and clinical trials) and knowledge translation via primary care electronic health record systems [6]. This involves the development of shared models and service infrastructure which allow for the efficient conduct of research. This is coupled with the delivery of decision support tools based on clinical evidence generated from the same electronic sources of primary care data.

2.1 Knowledge Representation in TRANSFoRm

A core element of TRANSFoRm is the development of shared models that allow for representation and exchange of the three distinct types of knowledge: research knowledge, routine healthcare knowledge and actionable knowledge. This relationship is shown in figure 1.

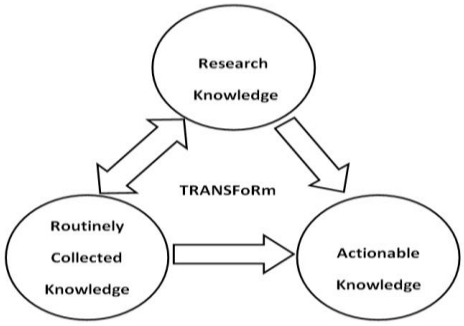

Fig. 1. – Conceptual relationship of Clinical Knowledge types in TRANSFoRm

From a decision support perspective the development of a model of actionable clinical knowledge is the core requirement. Actionable knowledge is knowledge that has been distilled from either research knowledge (generated from the conduct of controlled trials and epidemiological studies) and/or from routinely collected healthcare data. It is collected as part of consultations with patients and captured in electronic sources of patient data (such as electronic health records). This requires the application of data-mining and statistical analysis techniques to aggregated sources of electronic patient data to detect trends or patterns in the underlying data that may be used to infer diagnostic association rules. These are then used to construct computable clinical guidelines that can be deployed using decision support tools as part of clinical consultations with patients.

2.2 The Transition Project

The Transition project has demonstrated the feasibility of generating computable actionable knowledge from electronic sources of primary care patient data [7]. A more detailed description of work and methodology used as part of the Transition project has been described elsewhere [8].

The Transition project has utilized the International Classification of Primary Care, second edition (ICPC2) as a clinical classification to provide for the capture of patient data during consultations in family practice in four different countries [9]. The capture of the unambiguous clinical meaning of patient data as recorded in the electronic health record has been recognized as a requirement in the development of formal models of clinical knowledge [10]. A key conclusion of the Transition project was that not only was it feasible to generate actionable knowledge from coded sources of primary care patient data, but that the associations and calculated quantifications of primary care diagnostic cues to diagnostic outcomes were consistent across independent geographic regions.

The Transition project has captured patient data from four countries and quantified the association of ICPC2 coded diagnostic cues to specific diagnostic outcomes. This is presented as calculated likelihood ratios and confidence intervals within the context of a presenting patient reported reason for encounter and geographic region. An example subset of analysis of the association of the symptom ‘cough’ with ICPC2 coded outcomes in the context of the presenting reason for encounter ‘cough’ for patient data collected in the Netherlands is shown in figure 2. The strength of the association is categorized as ‘weak’, ‘strong’ or ‘not significant’ based on the relative value of the likelihood ratio and the width of the associated confidence interval (the methodology and associated calculations are fully described by the Transition Project [8]).

	The Netherlands	
Episode titles	LR+	LR-
Cough (R05)	**20.3 (19.9-20.7)**	*0.2 (0.2-0.2)*
Acute bronchitis/bronchiolitis (R78)	**16.2 (15.8-16.5)**	*0.3 (0.3-0.3)*
URTI head cold (R74)	**8.5 (8.2-8.7)**	0.6 (0.6-0.6)
Acute laryngitis/tracheitis (R77)	**12.5 (12.1-12.9)**	*0.4 (0.3-0.4)*
Sinusitis (R75)	*2.8 (2.6-3.0)*	0.9 (0.9-0.9)
Pneumonia (R81)	**8.5 (8.0-9.0)**	0.6 (0.5-0.6)
Influenza (R80)	*6.3 (5.8-6.7)*	0.7 (0.7-0.7)
Asthma (R96)	*8.0 (7.4-8.5)*	0.6 (0.6-0.6)
Other viral disease NOS (A77)	*2.5 (2.2-2.8)*	0.9 (0.9-0.9)
Whooping cough (R71)	**14.5 (13.7-15.3)**	*0.2 (0.2-0.3)*
Acute otitis media/myringitis (H71)	0.8 (0.7-1.0)	1.0 (1.0-1.0)
Symptoms/complaints throat (R21)	0.7 (0.5-0.8)	1.0 (1.0-1.0)
Tonsillitis (R76)	0.6 (0.5-0.8)	1.0 (1.0-1.0)
Adverse effect medication proper dose (A85)	*0.2 (0.2-0.3)*	1.1 (1.0-1.1)
Hayfever/allergic rhinitis (R97)	0.7 (0.6-0.9)	1.0 (1.0-1.0)
Symptoms/complaints chest (L04)	*0.3 (0.2-0.4)*	1.0 (1.0-1.1)
Hypertrophy tonsils/adenoids (R90)	1.7 (1.3-2.2)	1.0 (0.9-1.0)
Shortness of breath/dyspnea (R02)	0.9 (0.6-1.1)	1.0 (1.0-1.0)
Fever (A03)	0.8 (0.6-1.1)	1.0 (1.0-1.0)
COPD (R95)	*3.2 (2.5-4.2)*	0.9 (0.8-0.9)
General weakness/tiredness (A04)	*0.2 (0.1-0.2)*	1.1 (1.1-1.1)
Chronic bronchitis (R79/R91)	**9.8 (8.0-12.1)**	0.5 (0.4-0.6)

Black = not significant (LR+ <=2, LR- >=0.5, or wide CI)
Italics = weak predictor (LR+ >2-8, LR- 0.2-0.4, small CI)
Bold = strong predictor (LR+ >8, LR- <0.2, small CI)

Fig. 2. – Subset of Transition Project sample analysis for Netherlands showing calculated positive and negative likelihood ratios (LR) with associated confidence intervals (CI)

In the context of the TRANSFoRm project, the output of the analysis generated from the Transition project has provided one useful starting point for informing the modeling of diagnostic clinical evidence as a basis for future diagnostic decision support tool development.

3 Construction of an Ontological Model of Evidence

The usefulness and application of what can broadly be termed ‘semantic web’ technologies for modeling complex real-world systems has been demonstrated in a wide variety diverse settings including biomedicine, social-networking and on-line retailing [11]. The abstract representation of structures of hierarchical real-world concepts and the definition of the relationships that exist between them is addressed specifically in the area of ontology development. An ontology allows for the development of a portable, shareable, reusable abstract definition of the knowledge domain being modeled [12]. An example of this is the Basic Formal Ontology (BFO) developed as an upper ontology and reused by TRANSFoRm and many other diverse scientific settings [13].

From a clinical perspective, the use of named and bidirectional relationships between ontological clinical concepts enables querying of those concepts from a ‘top-down’ perspective or a ‘bottom-up’ perspective. This is useful in modeling of data that would be captured as part of the diagnostic workup process. We can work through our model from a top-level reason-for-encounter down to individual diagnostic cues and back up again in iterative cycles to work through potential differential diagnoses and investigate an individual diagnostic hypothesis [14].

3.1 Ontology Construction Methodology

Many formal methods have been proposed for the design, implementation and validation of ontologies [15-16]. The clinical evidence ontology will be used by a diagnostic decision support application allowing diagnostic workup of potential differential diagnoses to consider based on a presenting patient reason for encounter. A functional approach and definition of a decision support functional specification has driven ontology construction. The expression of functional requirements in the form of informal ontology 'competency questions' was selected as a suitable methodology. This allows formulation of competency questions and their expression using ontology query languages such as SPARQL for testing and validation of defined clinical scenarios.

3.2 Identification of Core Ontological Concepts

A review of the Transition project data identified core ontological concepts that need to be represented in the model. A subset of these showing Transition project definitions for the most important ones and associated examples is shown in table 1.

Table 1. Identified Core Ontological Concepts

General Concept Name	Description and Transition Data Example
Reason for Encounter	An agreed statement of the reason(s) why a person enters the health care system, representing the demand for care by that person. The reason for encounter should be recognized by the patient as an acceptable description of the demand for care. E.g. Cough (as a reason for encounter)
Diagnosis	Formal statement of the providers understanding of the patient's health problem, representing the establishment of an episode of care. It may be a symptom diagnosis or a disease diagnosis. E.g. Chronic Bronchitis
Diagnostic Cue	The symptoms, complaints, objective signs, and/or test results essential for labeling a health problem with a specific diagnosis. E.g. Cough (as a symptom)
Quantification	A quantifiable measure of the association of a diagnostic cue to the presence or absence of a particular diagnosis. E.g. A calculated likelihood ratio value (positive or negative) and associated confidence intervals
Evidence Population	A concept capturing the demographic or population characteristics from which a particular quantification was obtained. E.g. Sex, age, ethnicity or country

3.3 Construction, Population and Hosting Model of Evidence

An ontology of clinical evidence has been constructed for TRANSFoRm using Protégé version 4.1 based on Web Ontology Language (OWL) and Resource Description Framework Schema (RDFS) ontology languages [17-18]. In order to support future development of the decision support tool and to allow for dynamic population of ontology data from analysis done on electronic sources of patient data, the ontology has been deployed to and hosted using the Sesame platform [19]. This provides an open source triple-store backend that has compared favorably in performance tests with other available solutions [20]. It also provides a platform for development and testing of ontology queries using Simple Protocol and RDF Query Language (SPARQL) queries to test the conceptual completeness of the ontology design and the accuracy of generated results. The Transition analysis data for the symptom 'cough' was then manually populated into the ontology.

3.4 Testing and Validating the Model of Evidence

Informal competency questions were translated to formal SPARQL queries to test that all required clinical questions could be expressed using the ontology ensuring conceptual completeness. All generated outputs to those queries were checked for consistency with respect to the original Transition data that was modeled. A sample clinical competency question is: identify the diagnoses for which the symptom X is a strong predictor in the population Y? The formal equivalent query constructed to test for the symptom instance 'cough' in the context of the population instance 'Netherlands' and the associated test result is shown in table 2.

Table 2. Sample SPARQL Formal Query and Results

Formal SPARQL Query	Result (Concept Instances)

`SELECT ?anyDiagnosis`
`WHERE {Cough hasQuantification ?anyQuantification.`
`?anyQuantification hasPosLREvidenceStrength "Strong predictor"^^xsd:string.`
`?anyQuantification hasEvidenceCountry Netherlands.`
`?anyQuantification hasQuantificationDiagnosis ?anyDiagnosis.}`
	`Cough`
	`AcuteBronchitis`
	`URTIHeadCold`
	`AcuteLaryngitis`
	`Pneumonia`
	`WhoopingCough`
	`ChronicBronchitis`

The query result is correct with respect to the original Transition project data shown previously in figure 2. The characteristics associated with these results could be investigated further using additional SPARQL queries based on the ontology concepts and relationships. The complete list of clinical competency questions developed was successfully translated into equivalent formal SPARQL queries and tested against the host platform to ensure conceptual validity and accuracy of results against the original Transition project data.

4 Future Work

The work done to date has focused on development of a back-end model of evidence and a hosting platform. Initial work is now starting on building a web based clinical evidence service application around this. The web service will support two major interfaces: a query interface for asking diagnostic clinical questions to the web service, and an update interface to allow for regular update of the ontology evidence as generated from data mining and analysis modules applied to aggregated sources of primary care data such as the Transition project. The final stage of work will involve the development of the actual decision support tool. This tool will be integrated with a primary care EHR system to be triggered based on the reason for encounter to collect ontologically controlled diagnostic cues.

5 Conclusions

The ontology models of general evidence developed as part of TRANSFoRm were conceptually descriptive enough to model the ICPC2 based data analysis of the diagnostic associations with the symptom 'cough' in the context of four separate population regions. By carrying out additional data mining and analysis on more diagnostic cues it is feasible to develop a full picture of ICPC2 coded diagnostic cues and their associations that have been also been quantified using likelihood ratios based on the underlying patient data that is population specific. The Sesame platform provides a suitable ontology hosting mechanism that TRANSFoRm will utilize to develop a back end web based evidence service to provide decision support. This will be based on evidence generated from electronic sources of primary care data that will be populated or changed dynamically as that underlying patient data grows or changes. This is consistent with the goal of implementing translational and evidence based decision support based on the electronic health record.

References

1. Sackett, D.L., Rosenberg, W.M., Gray, J.A., Haynes, R.B., Richardson, W.S.: Evidence based medicine: what it is and what it isn't. BMJ. 312, 71–72 (1996).
2. Greenhalgh, T.: How to Read a Paper: The Basics of Evidence-Based Medicine. John Wiley & Sons (2010).

3. Haines, A., Jones, R.: Implementing findings of research. BMJ. 308, 1488–1492 (1994).
4. Grol, R., Grimshaw, J.: From best evidence to best practice: effective implementation of change in patients' care. Lancet. 362, 1225–1230 (2003).
5. Lang, E., Wyer, P., Haynes, B.: Knowledge Translation: Closing the Evidence-to-Practice Gap. Annals of Emergency Medicine. 49, 355–363 (2007).
6. TRANSFoRm, http://www.transformproject.eu
7. The Transition Project, http://www.transitieproject.nl
8. Soler, J.K., Okkes, I., Oskam, S., van Boven, K., Zivotic, P., Jevtic, M., Dobbs, F., Lamberts, H., for the Transition Project: The interpretation of the reasons for encounter 'cough' and 'sadness' in four international family medicine populations. Informatics in Primary Care. In press (2012).
9. WHO International Classification of Primary Care, Second edition (ICPC-2), http://www.who.int/classifications/icd/adaptations/icpc2/en/index.html.
10. Coiera, E.: Guide to Health Informatics. Hodder Arnold Publishers (2003).
11. Kashyap, V., Bussler, C., Moran, M.: The semantic web: semantics for data and services on the web. Springer-Verlag New York Inc (2008).
12. Allemang, D., Hendler, J.: Semantic Web for the Working Ontologist, Second Edition: Effective Modeling in RDFS and OWL. Morgan Kaufmann (2011).
13. Basic Formal Ontology homepage, http://www.ifomis.org/bfo/home
14. Kostopoulou, O.: Diagnostic Errors: Psychological Theories and Research Implications. In: Arts, B.H. of M. and the and Development, A.S. of P.C.R. and (eds.) Health Care Errors and Patient Safety. pp. 95–111. Wiley-Blackwell (2009).
15. Grüninger, M., Fox, M.S.: Methodology for the Design and Evaluation of Ontologies. In: Workshop on Basic Ontological Issues in Knowledge Sharing. Montreal (1995).
16. Fernandez-Lopez, M., Gomez-Perez, A.: Overview and analysis of methodologies for building ontologies. The Knowledge Engineering Review. 17, 129–156 (2002).
17. The Protégé Ontology Editor and Knowledge Acquisition System, http://protege.stanford.edu
18. OWL 2 Web Ontology Language Document Overview, http://www.w3.org/TR/2009/REC-owl2-overview-20091027
19. openRDF Sesame homepage, http://www.openrdf.org
20. Bioontology.Org, http://www.bioontology.org/wiki/images/6/6a/Triple_Stores.pdf