1-1-2016

An Electronic Evidence Base Supporting Derivation, Dissemination and Learning for Diagnostic Clinical Prediction Rules in Primary Care Practice

Derek T. Corrigan
Royal College of Surgeons in Ireland, derekcorrigan@rcsi.ie

Citation
An Electronic Evidence Base Supporting Derivation, Dissemination and Learning for Diagnostic Clinical Prediction Rules in Primary Care Practice

Derek T Corrigan BSc MSc
Department of General Practice
RCSI

A thesis submitted to the School of Postgraduate Studies, Faculty of Medicine and Health Sciences, Royal College of Surgeons in Ireland, in fulfilment of the degree of
Doctor of Philosophy

Supervisors:
Professor Tom Fahey (RCSI)
Dr. Damon Berry (Dublin Institute of Technology)

April 2016
I declare that this thesis, which I submit to RCSI for examination in consideration of the award of a higher degree Doctor of Philosophy is my own personal effort. Where any of the content presented is the result of input or data from a related collaborative research programme this is duly acknowledged in the text such that it is possible to ascertain how much of the work is my own. I have not already obtained a degree in RCSI or elsewhere on the basis of this work. Furthermore, I took reasonable care to ensure that the work is original, and, to the best of my knowledge, does not breach copyright law, and has not been taken from other sources except where such work has been cited and acknowledged within the text.

Signed

_______________________________________

Student Number

_______________________________________

Date

_______________________________________
# Table of Contents

1  Introduction .................................................................................................................. 23

1.1 CPRs as a Form of Evidence Based Medicine ......................................................... 25

1.2 The Computable CPR as Part of a Learning Health System ................................. 27

1.3 Research Questions ...................................................................................................... 29

1.4 Aims/ Objectives of this Research ............................................................................. 32

1.5 Summary of Methods and Thesis Structure ............................................................ 35

2  State of the Art – CPR Development in Primary Care .............................................. 37

2.1 Objectives of Clinical Literature Review .................................................................. 37

2.2 What do we Mean by the Term ‘Primary Care’ ......................................................... 37

2.3 Causes and Implications of Error in Primary Care .................................................... 38

2.3.1 What is an ‘Adverse Event’? .................................................................................. 39

2.3.2 Why do Adverse Events Happen? ......................................................................... 40

2.3.3 When do Adverse Events Happen? ...................................................................... 43

2.3.4 When Diagnosis in Primary Care Goes Wrong ..................................................... 46

2.4 Cognitive Strategies Used for Diagnostic Decision Making ...................................... 48

2.4.1 Intuition and Heuristics as Part of the Diagnostic Process ..................................... 49

2.4.2 Cognitive Bias in Heuristics .................................................................................. 49

2.5 Improving Diagnostic Decision Making through Practice of Evidence Based Medicine .......................................................................................................................... 51

2.5.1 Evidence Based Medicine Evolves ........................................................................ 52

2.6 What are CPRs and How are they Applied? ............................................................... 53

2.6.1 Types of CPR ......................................................................................................... 53

2.6.2 CPRs as a Form of Probabilistic Diagnostic Reasoning ....................................... 53

2.6.3 CPRs as a Form of Threshold Based Risk Interpretation and Action 54

2.6.4 The Formal Structure of a CPR ............................................................................ 55
3.7 Data mining eCPRs from the EHR ............................................................... 86
  3.7.1 Observation, Measurement and Evidence Discovery - the
  Generic Data Mining Process ................................................................. 86
  3.7.2 Structured and Unstructured Data Mining ...................................... 87
  3.7.3 Data Mining for eCPRs and Clinical Decision Support ............... 88
3.8 Decision Support Tools as a ‘Medical Device’ ..................................... 89
  3.8.1 EU Definitions .................................................................................. 90
  3.8.2 EU Medical Device Classes ............................................................. 91
  3.8.3 Compliance Procedure ................................................................... 92
  3.8.4 EU Applicable Standards ................................................................. 93
  3.8.5 Medical Devices Outside the EU ..................................................... 93
  3.8.6 Research Interpretation of Medical Device Directive .................... 94
3.9 Technical Literature Review Conclusions ............................................ 95
4 Conceptual Framework for Development of eCPRs ................................ 98
  4.1 Developing Smarter eCPRs as Part of the LHS ................................. 98
  4.2 A Working Definition of Interoperability for ICT Systems ............... 99
    4.2.1 Syntax .......................................................................................... 100
    4.2.2 Semantics .................................................................................... 101
    4.2.3 Transport and Services ................................................................. 101
  4.3 Components of the eCPR Maturity Model .......................................... 102
    4.3.1 Level 1: Literature Based CPRs ................................................... 102
    4.3.2 Level 2: Electronic Document Based eCPRs .............................. 103
    4.3.3 Level 3: Electronic Computable Individual eCPR Tool ............... 104
    4.3.4 Level 4: Service-oriented Generalised eCPR .............................. 106
    4.3.5 Level 5: eCPRs with Terminology Services Integration ............. 109
    4.3.6 Level 6: Learning, Versionable eCPR ......................................... 110
5 eCPR Framework Functional Design, Implementation and Results 112
5.1 Assumptions and Scope Related to Implementation

5.2 TRANSFoRm Acknowledgement and Work Delineation

5.3 Level 1 - Selection of Working Clinical Use Cases

5.3.1 The Alvarado Score as a CPR Example

5.4 Level 4: Generalised Models of Clinical Evidence

5.4.1 Ontology Derivation and Validation Based on Competency Questions

5.4.2 Definition of Functional Requirements as Clinical Competency Questions

5.4.3 General Model of Evidence Concepts

5.4.4 eCPR Model of Evidence Concepts

5.5 Ontology Model Construction - Development Tools

5.5.1 Modelling Multi-morbidity

5.5.2 Evidence Service Construction and Architecture

5.5.3 Ontology Triple Store Selection and Performance Considerations

5.5.4 Development of SPARQL Queries

5.5.5 Sesame Application Programming Interface (API)

5.5.6 Ontology Model Validation

5.5.7 Model Testing

5.6 Level 4: Service Oriented Provision of Evidence Content

5.6.1 Jersey Java implementation of REST Services

5.6.2 Clinical Evidence Service Interfaces and Methods

5.6.3 Sesame REST API Support for Ad-hoc Queries

5.7 Level 5: CPRs with Terminology Services Integration

5.8 Level 6: Learning, Versionable Electronic CPRs

5.8.1 Data Mining and Evidence Population for UTI and Pyelonephritis

5.8.2 Clinical Association Rule Review

5.8.3 Rule Review and Curation in Practice
5.8.4 Association Rule Export ...................................................... 162
5.8.5 Ontology Rule Import and Update ........................................ 163
5.8.6 Rule Availability Through the Web Service .............................. 165
5.8.7 Comparisons with the Literature ........................................... 166
5.8.8 Comparison between Populations .......................................... 168
5.8.9 eCPR Workbench Specification and Development .................... 169

5.9 Workflow Integration Formalising the Diagnostic Process ............ 173
5.9.1 ‘Suggesting’ Phase Diagnostic Strategies ............................. 178
5.9.2 ‘Alerting’ Phase Diagnostic Strategies (254) .......................... 179

5.10 Evidence Service Evaluation in Practice .................................. 180
5.10.1 Development of Diagnostic Tools Using the Clinical Evidence
Service ................................................................................... 180
5.10.2 Evidence Service Integration with TRANSFoRm Diagnostic
Decision Support Tool ............................................................... 182
5.10.3 Evaluation Based on Review of Desirable Features for Decision
Support Systems ..................................................................... 185
5.10.4 Interpretation of Architecture Evaluation ............................... 190

6 Discussion / Conclusion ................................................................ 191
6.1 Delivery of Research Question Implementation Aims ..................... 191
6.2 Delivery of Research Question Conceptual Aims ............................ 194
6.3 Limitations of this Work ............................................................. 197
6.3.1 Bias and Error in Underlying Data Capture and Coding ............ 197
6.3.2 Granularity of Coding Scheme Used ..................................... 200
6.3.3 Limited Literature Based Evaluation of the Data Mined Evidence 202
6.3.4 A Limited eCPR Learning Cycle Without a Feedback Loop ....... 203
6.3.5 Constraints on Underlying Terminology Unable to Facilitate
Epidemiological ‘Unknowns’ ....................................................... 204
List of Peer-Reviewed Publications

**Corrigan D.** An ontology driven clinical evidence service providing diagnostic decision support in family practice. Proceedings of the AMIA Summit on Clinical Research Informatics; San Francisco, USA 2015.


List of Oral Presentations/ Conference Proceedings


Invited Presentations


Corrigan D, Curcin V. Knowledge Modeling for Diagnostic Support. ENJECT Training school on the Learning Health System, Dublin, Ireland, December 14th 2015.
**List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BFO</td>
<td>Basic Formal Ontology</td>
</tr>
<tr>
<td>CA</td>
<td>Competent Authority</td>
</tr>
<tr>
<td>CDIM</td>
<td>Clinical Data Integration Model</td>
</tr>
<tr>
<td>CDSS</td>
<td>Clinical Decision Support System</td>
</tr>
<tr>
<td>CPR</td>
<td>Clinical Prediction Rule</td>
</tr>
<tr>
<td>DLL</td>
<td>Dynamic Linked Library</td>
</tr>
<tr>
<td>eCPR</td>
<td>Electronic Clinical Prediction Rule</td>
</tr>
<tr>
<td>EHR</td>
<td>Electronic Health Record</td>
</tr>
<tr>
<td>EMR</td>
<td>Electronic Medical Record</td>
</tr>
<tr>
<td>EPR</td>
<td>Electronic Patient Record</td>
</tr>
<tr>
<td>ICD10</td>
<td>International Classification of Diseases Version 10</td>
</tr>
<tr>
<td>ICPC2</td>
<td>International Classification of Primary Care Version 2</td>
</tr>
<tr>
<td>JSON</td>
<td>Javascript Object Notation</td>
</tr>
<tr>
<td>KNIME</td>
<td>The Konstanz Information Miner</td>
</tr>
<tr>
<td>NB</td>
<td>Notified Body</td>
</tr>
<tr>
<td>OWL</td>
<td>Web Ontology Language</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomised control trial</td>
</tr>
<tr>
<td>RDF</td>
<td>Resource Description Framework</td>
</tr>
<tr>
<td>REST</td>
<td>Representational state transfer</td>
</tr>
<tr>
<td>RFE</td>
<td>Reason for Encounter</td>
</tr>
<tr>
<td>SPARQL</td>
<td>SPARQL Protocol and RDF Query Language</td>
</tr>
<tr>
<td>SAML</td>
<td>Security Assertion Markup Language</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>SNOMED</td>
<td>The Systematized Nomenclature of Medicine</td>
</tr>
<tr>
<td>TRIPOD</td>
<td>Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis</td>
</tr>
<tr>
<td>UMLS</td>
<td>Unified Medical Language System</td>
</tr>
<tr>
<td>UTI</td>
<td>Urinary Tract Infection</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
</tbody>
</table>
List of Figures

Figure 2-1 Diagnostic stages and strategies employed in primary care – courtesy of Heneghan et al.2009 (51).................................................................44
Figure 2-2 Strategies used by general practitioners in making an initial diagnosis - courtesy of Heneghan et al. 2009 (51)........................................45
Figure 2-3 Strategies used by general practitioners in refining a diagnosis - courtesy of Heneghan et al. 2009 (51).........................................................46
Figure 2-4 CRICO analysis of where errors were made in diagnostic assessment of 2,596 malpractice claims, CRICO 2015 (47)..........................47
Figure 2-5 CPR development lifecycle stages ..................................................59
Figure 3-1 Key phases in the evolution of clinical decision support systems - Courtesy of Wright et al. (122).................................................................74
Figure 4-1 A Multi-step maturity model for the implementation eCPRs........102
Figure 4-2 The HRB Centre for Primary Care Research eCPR register ........104
Figure 5-1 Architecture of an eCPR based learning health system.................114
Figure 5-2 The Alvarado Score for assessment of likelihood of appendicitis (courtesy of Alvarado 1986) (232).................................................................119
Figure 5-3 Core general model of diagnostic evidence ontology concepts and relations. ..................................................................................................127
Figure 5-4 CPR ontology model concepts and relationships .......................128
Figure 5-5 Defining ontology concepts in Protégé ...........................................130
Figure 5-6 Defining ontology relationships in Protégé.....................................131
Figure 5-7 Defining ontology class diagnosis instances in Protégé ...............132
Figure 5-8 Models of morbidity.......................................................................133
Figure 5-9 Architecture implementation technologies ..................................135
Figure 5-10 Clinical evidence architecture overview .......................................136
Figure 5-11 Sesame triple store workbench browser ....................................137
Figure 5-12 Sesame SPARQL query testing ..................................................139
Figure 5-13 An evidence service reply describing differential diagnoses to consider related to a presenting reason for encounter of abdominal pain......144
Figure 5-14 An evidence service reply describing a symptom collection for Urinary Tract Infection..............................................................145
Figure 5-15 Core general model of diagnostic evidence ontology concepts and relations with additional support for semantic concepts. .................................147
Figure 5-16 Example of cue ontology concept instance for ‘history of irritable bowel syndrome’ with associated code and synonym.................................148
Figure 5-17 Evidence service reply describing a symptom collection for Urinary Tract Infection with associated read code bindings.................................149
Figure 5-18 TransHIS episode of care structure ........................................151
Figure 5-19 Data mined association rule structure ....................................151
Figure 5-20 TransHIS data mining process .............................................153
Figure 5-21 Web rule viewer selection of rules filters .................................154
Figure 5-22 The main rule viewer details screen ......................................154
Figure 5-23 Generation of association rules export XML from the web rule sender ..................................................................................................162
Figure 5-24 Flowchart for ontology update of evidence from XML export rule files .............................................................................................164
Figure 5-25 REST evidence output of imported quantified evidence rules .....166
Figure 5-26 eCPR construction tool to create electronic CPRs from quantified evidence obtained through a data mining process .........................169
Figure 5-27 A Web service call to a data mined eCPR (example of Level 6 learning, versionable eCPR) .................................................................172
Figure 5-28 TRANSFoRm 2 phase diagnostic process ..............................174
Figure 5-29 TRANSFoRm diagnostic workflow and interaction of decision support components ..............................................................................177
Figure 5-30 An XML patient evidence case submitted to the evidence service for a female patient presenting with chest pain and symptoms including fatigue. ................................................................................................181
Figure 5-31 The integrated diagnostic decision support window accessible from the patient EHR record shown in the background ............................183
List of tables

Table 1-1 Research question 1 knowledge gaps and aims........................................33
Table 1-2 Research question 2 knowledge gaps and aims........................................34
Table 1-3 Research question 3 knowledge gaps and aims........................................35
Table 3-1 Example of triples and inverse relationships in the diagnostic domain
...........................................................................................................................................81
Table 5-1 Clinical competency questions to describe requirements for
developing the ontology models.....................................................................................123
Table 5-2 Summary of CPR ontology concepts and instances.................................141
Table 5-3 Clinical competency questions implemented as formal ontology
queries with results..............................................................................................................142
Table 5-4 Positive likelihood ratios for associated RFEs (label and ICPC code
listed) and the episode title “UTI” in two populations.................................................156
Table 5-5 Positive likelihood ratios for associated RFEs (label and ICPC code
listed) and the episode title “pyelonephritis” in two populations............................159
Table 5-6 Data mined eCPR supporting diagnosis of Urinary Tract Infection in
primary care......................................................................................................................171
Table 5-7 Data mined eCPR supporting diagnosis of Pyleonephritis in primary
care..................................................................................................................................171
Table 5-8 Desirable functionality for implementing clinical decision support
systems amended from Wright et al.(128) ....................................................................185
Table 6-1 Implementation aims of research question 1..............................................193
Table 6-2 Implementation aims of research question 2..............................................193
Table 6-3 Implementation aims of research question 3..............................................194
Table 6-4 Conceptual aims of research question 1......................................................196
Table 6-5 Gaps in knowledge regarding use of CPRs in practice............................212
Abstract

Diagnostic error is a threat to patient safety in the context of primary care. Clinical prediction rules (CPRs) are a form of structured evidence based guideline that aim to assist clinical reasoning through the application of empirically quantified evidence to evaluate patient cases. Their acceptance in clinical practice has been hindered by literature-based dissemination and doubts regarding their wider applicability. The use of CPRs as part of electronic decision support tools has also lacked acceptance for many reasons: poor integration with electronic health records and clinician workflow, generalised guidelines lacking patient-specific recommendations at point-of-care, static rule based evidence that lacks transparency and use of proprietary technical standards hindering interoperability.

The ‘learning health system’ (LHS) describes a distributed technology based infrastructure to generate computable clinical evidence and efficiently disseminate it into clinical practice. This research describes an LHS based on computable CPRs for diagnostic decision support that makes use of aggregated sources of primary care electronic health record data to derive and disseminate computable CPRs.

Based on a literature review of clinical and technical best practice regarding use of CPRs, a theoretical model for CPRs supporting two critical aspects for a successful LHS is proposed: the model representation and translation of clinical evidence into effective practice, and the generation of curated clinical evidence that can dynamically populate those models thus closing the learning health system loop. A functional implementation of the theoretical model demonstrates an infrastructure that is model-driven, service oriented, constructed using open standards, and supports a learning evidence base derived from electronic sources of patient data.

A number of challenges exist for the LHS community to consider including medico-legal responsibility for generated diagnostic evidence, developing trust in the LHS, constraints imposed by clinical terminologies on evidence generation, and quality and bias of underlying EHR data for evidence generation.
Acknowledgements

I would like to first acknowledge the truly excellent clinical and technical supervision provided to me during this work by Tom Fahey and Damon Berry respectively. They have only been a source of encouragement and good advice whenever I have needed it and made this initially daunting task seem possible at all times.

I would like to also thank the many truly excellent people I have been privileged to work with as part of the TRANSFoRm project. Without these people I don’t think I could ever have understood ‘the bigger picture’. In particular I would like to thank:

- Gary Munnelly, Lucy Hederman (Trinity College Dublin) for their excellent work helping to implement the clinical evidence service
- Dr. Jean Karl Soler (Mediterranean Institute for Primary Care Research), Professor Przemyslaw Kazienko, Tomasz Kajdanowicz (Wroclaw University of Technology) , Roxana Danger Mercaderes (Imperial College London) for their help in developing and implementing the data mining methodology for the clinical evidence service.

I am heavily indebted to the TRANSFoRm management team of Professor Brendan Delaney (Imperial College London) and Doctor Vasa Curcin (Kings College London). They showed immense patience at all times during the TRANSFoRm project but particularly when I was starting out and needed it most.

This work was partially funded by the Health Research Board (HRB) of Ireland through the HRB Centre for Primary Care Research under Grant HRC/2007/1.

TRANSFoRm is partially funded by the European Commission – DG INFSO (FP7 247787).
Dedication

There are always people who, though not directly involved, are essential to any possibility of your work being successful. They help you contextualise and appreciate the most important things in life. The birth of this research has broadly ‘aligned’ with the birth of my daughter Siún. Received wisdom will tell you that starting a family and commencing a PhD. at the same time is a guaranteed recipe for disaster. I would strongly dispute that. There is nothing more inspiring than a wife who teaches your three and half year old daughter to ask her daddy for fun how his ‘doctor of philosophy’ is going. If that doesn’t motivate you then I’m afraid nothing will. Éimear and Siún, I love you, thank you and salute you.
1 Introduction

‘Good sense is, of all things among men, the most equally distributed; for everyone thinks himself so abundantly provided with it, that those even who are the most difficult to satisfy in everything else, do not usually desire a larger measure of this quality than they already possess’ (1)

Although we are all aware of it, we do not like to think too deeply upon our own capacity to make mistakes that may lead to incorrect decisions. This is particularly the case when those decisions are related to our working lives and have the potential to result in severe consequences for other people in our care. Diagnostic error is a major threat to patient safety, particularly in the context of the primary care clinical setting (2, 3). A study by Sanders estimates the number of diagnostic errors in primary care as being between 5 and 80 in 100,000 consultations (4). Other studies have estimated higher rates of diagnostic error. A systematic review by Berner concluded that ‘in aggregate, studies consistently demonstrate a rate of diagnostic error that ranges from <5% in the perceptual specialties (pathology, radiology, dermatology) up to 10% to 15% in most other fields’ (5).

The application of evidence-based medicine (EBM) has long been advocated as one way of ensuring standards of consistency in the application of diagnostic assessment with a view to improving clinical practice more generally. To the great frustration of many clinical staff, one of the biggest challenges for the effective practice of evidence-based medicine is that the corpus of clinical knowledge has largely been a static, document-based but ever more rapidly growing evidence base (6).

The evidence base relating to diagnosis is frequently provided in the form of national repositories of document-based guidelines which may be different depending on their applicability to different demographic populations (7). The ability of clinicians to efficiently search through these large bodies of on-line guidance, identifying the most suitable evidence to apply in any particular presenting patient context, has become an assumed technical skill for the effective dissemination and application of evidence-based guidelines in practice.
The need to efficiently organise, curate and deploy up-to-date clinical knowledge for use in clinical practice has been demonstrated by the increasing role played by information and communications technology (ICT) in filtering and identifying literature for both research and clinical practice in the form of internet based search engines (9).

This traditional paradigm of evidence gathering based on searching ‘generalised guidelines’ has begun to change with the wider adoption of patient records held in digital formats rather than in written patient charts. These digital records may be in the form of shared repositories of patient data, directly accessible by multiple users across wider health system functions (referred to as electronic health records (EHRs)) or patient records that are accessible in isolation within a single primary care practice (electronic patient records (EPRs)) (10, 11). For the purposes of clarity in this work, the shared access distinction is not critical and we will only refer to EHRs as the generic term covering both from now on.

The new evidence gathering paradigm refers to an automated process that uses the EHR to drive triggering, collection and application of patient-specific clinical guidance relevant to any presenting patient case. ICT supports clinical practice by enabling staff, through provision of EHR integrated tools, to contextually access the latest evidence-based and research trialled guidelines, assisting clinicians with patient assessment at the point of care (12).

This enhanced EHR functionality will in effect turn EHRs into EHR driven clinical decision support systems (CDSSs); ‘computer systems that are designed to impact clinician decision making about individual patients at the point in time that these decisions are made’ (13). The ultimate vision is articulated by president Barack Obama in the recently announced US ‘precision medicine’ initiative, where clinical guidance and associated care is carried out based on evidence gathering and interpretation appropriate to the unique clinical context and characteristics of any individual presenting patient case, rather than using ‘boiler-plate’ generic clinical guidelines (14).

Significant progress has been made in the development of clinical decision support systems (CDSS) with varying degrees of success depending on the
clinical area being addressed (15). A fundamental problem arises where current guidelines may only exist in paper rather than electronic formats. Current paper-based guidelines are not easily translated to a machine interpretable formats that computer-based systems can interpret or reason with. This is reflected in the sometimes radically different, and competing approaches of different standards available for representing clinical data for use by ICT systems such as electronic health records (EHR) (16).

Paper-based guidelines are problematic for a number of reasons. The traditional structure of many clinical guidelines is to present information relating to single conditions in isolation without cross-referencing other potentially relevant information. Multi-morbidity however is a reality for many patients, particularly those with chronic conditions, and they present in primary care practice with more than one clinical condition at the same time (17, 18). The implications of following guidelines in isolation include increased risk of polypharmacy, driving overtreatment in the form of inappropriate medication use that can lead to increased risk of adverse drug reactions (19, 20). Recommendations to address this problem include using ICT based approaches to present guidelines in a more cross-referenced and patient focussed fashion use electronic patient data as a contextual trigger (21).

Paper-based guidelines are also static in nature and cannot automatically be combined or transformed into derived clinical knowledge as new clinical research data develops and becomes available. In contrast, electronically accessible sources of clinical knowledge provide an opportunity to do further data-analysis on the underlying data sources using the application of statistical or data-mining algorithms. As will be shown later, this data-mining process can be used to periodically ‘discover’ new clinical knowledge and ‘reassess’ the evidence baseline. A fundamental challenge therefore is how to represent known clinical facts or knowledge in a format that can be used and easily shared by different computer systems.

1.1 CPRs as a Form of Evidence Based Medicine
Definitions of what constitutes the practice of evidence based medicine have subtly changed over time and will be discussed in detail in section 2.5.1. Initial
definitions focussed on a contextual process to systematically search and review available evidence for any presenting clinical problem giving due consideration to the clinical characteristics of each individual patient presenting to practice (22). This provides a rigorous alternative to reliance on more heuristic methods that evaluate the evidence at hand in a more subjective and general way that is based solely on ‘clinical experience’.

More recent definitions have emphasised the need to mathematically quantify the ‘best available’ evidence. These definitions advocate the use of formal risk based models to quantify evidence obtained from representative patient populations conducted in the form of observational studies (8). The clinical guidelines that are consistent with these more data-driven definitions are well structured and derived based on mathematical models that empirically quantify the contribution of clinical variables to a particular clinical outcome.

One such example that is the focus of this research and described fully in section 2.6 is the development of a structured clinical guideline known as a ‘clinical prediction rule’ (CPR). CPRs have a well-defined development lifecycle and are widely reported in research literature in a number of clinical domains (23). CPRs may be used to assist diagnosis through assessing the risk of a patient having a particular clinical condition. Or CPRs may be used to assess the likely prognosis for a patient already presenting with a diagnosed condition. The focus of this research is on diagnostic CPRs only and prognostic CPRs will not be considered. Diagnostic CPRs have been shown to be a useful clinical tool that may be used as part of a broader diagnostic strategy to assist with clinical decision making, although their acceptance and widespread adoption in frontline clinical practice has been limited for reasons that shall be discussed in detail in section 2.6 (24).

It has been suggested that one way of promoting the effective use of CPRs is to deploy them in electronic formats that can be incorporated into CDSSs that support clinicians to more easily identify suitable CPRs to apply for any presenting patient case (25).
1.2 The Computable CPR as Part of a Learning Health System

There have been some notable attempts described in literature to develop fully computable CPRs that are integrated and triggered from patient records provided by an EHR with encouraging results (12, 26, 27). To date these attempts have largely focussed on implementing proprietary models of limited numbers of static CPRs that can be integrated into the IT infrastructure of particular clinical institutions (12). These approaches do not provide for a more generalisable and widely accessible CPR model. This would define representation of the highly structured CPR in a generic form that could be used to represent any chosen CPR, enabling integration with any chosen third party consumer using openly available technology standards. An opportunity therefore exists to develop models of dynamically generated CPRs based on openly available technical standards and a service oriented approach that enables integration with any desired third party consumer.

The computable representation and integration of CPRs with EHRs is only one part of what can be described as an overall translational solution. It addresses the need to more easily disseminate CPR knowledge into frontline clinical practice. Another related aspect is the actual derivation and generation of CPR knowledge itself, through what could be termed a knowledge generation process.

Manually populating and maintaining CPRs obtained from literature into models for electronic dissemination is not a feasible solution in the long-term. Such a process involves the update of clinical evidence to take into account the ever changing body of literature based evidence along with the large numbers of clinical codes and terminologies that are required to represent them. This represents a manually intensive, time consuming and error-prone task. The ultimate translational goal is to develop CPRs that can be derived from quantified data extracted from aggregated sources of EHR patient data. A relatively recent development that attempts to define the requirements for such systems to exploit the value of EHR data is the emerging definitions of what is termed the Learning Health System (28, 29). The LHS presents an ambitious vision of an ICT infrastructure that can generate valid clinical knowledge.
generally, curate and manage that data, and easily disseminate it into clinical practice.

A key distinction should be made at this point relating to the knowledge generation process as described by the LHS. The LHS describes the integration and management of potentially three distinct types of health knowledge:

- routinely collected health knowledge (e.g. EHRs)
- research knowledge in the form of electronically conducted research studies and observational cohort studies from electronic registries
- derived clinical knowledge through the application of data mining and evidence discovery techniques applied to both routine and research study data (30).

The ability to conduct research studies electronically is a complex research topic in itself. This has been specifically addressed by European research projects such as TRANSFoRm and EHR4CR (31, 32). These projects aim to provide an electronic infrastructure to support the formal process of conducting well-structured research studies electronically. The evidence generation process as described in this work however will solely focus on the use of aggregated sources of routinely collected healthcare knowledge in the form of EHRs. The use and conduct of structured research studies for evidence generation will not be investigated further here but is acknowledged as important in the wider context.

Three key research questions are therefore addressed relating to the development of electronic and computable CPRs (referred to from now on as ‘eCPRs’) and will be framed as a specific instance of the more general LHS approach. It is also useful at this point to prevent any ambiguity regarding future use of the terms CPR and eCPR. The term CPR will be used to refer to clinical prediction rules as derived and used in the current traditional literature-based formats. The term eCPRs will refer to electronic clinical prediction rules as described in this research, disseminated in a computable format and capable of integration with third party ICT tools.
1.3 Research Questions

As will be discussed and shown in detail in chapter 2, a review of the relevant literature has highlighted a developing interest in research around derivation and application of CPRs. This research can be broadly defined and looked at from both a clinical and a technical perspective. The technical perspective suggests potential solutions to address limitations inherent in the traditional clinical perspective on use of CPRs. The literature review in chapter 2 is structured in two sections to reflect these distinct research perspectives on CPRs.

Research into the application of traditional literature described CPRs has demonstrated their efficacy in both general practice and secondary care whilst highlighting a number of barriers and limitations to their wider acceptance and use (25, 33, 34). A study by McGinn for example that trialled two well-known CPRs as part of an ICT based intervention reduced unnecessary antibiotic prescribing and lab testing with associated cost reductions. The absolute antibiotic prescribing rate in pneumonia patients was reduced by 9.2%. The absolute rate of streptococcal testing for pharyngitis patients was reduced by 12.1% (12).

Barriers to successful CPR development have been identified around the need for broader validation, impact analysis and methods of dissemination in clinical practice. This has resulted in emerging research into the application of ICT systems as a potential solution to these problems with a view to integrating computable versions of CPRs triggered by electronic health records to provide clinical decision support systems. This emerging understanding of the applications of ICT to develop computable CPRs highlights a significant gap that exists in the literature.

The ICT approaches to date typically have focussed on very specific instances of CPRs addressing particular presenting patient problems. These approaches have embedded technical implementations of specific CPRs as single applications within a proprietary institution health system using defined coding schemes. Whilst these approaches have demonstrated promise within their own silos of implementation, the wider generalisability and interoperability of these
approaches in practice is open to question. An opportunity therefore exists to propose a more generalisable solution in the form of a service oriented approach, detaching the evidence base from the application itself, to represent any CPR as a form of computable evidence making full use of suitable open technical standards.

This generalisable approach can be related to wider research initiatives beyond CPRs. These wider research developments are happening in parallel and have potential to add significant value to this proposal. Significant research has been conducted in the last five years around possibilities for developing a ‘Learning Health System’ that utilises aggregated sources of electronic health record data to generate clinical knowledge. Much of this material has been conceptual in nature focussing on defining what the core elements of a Learning Health System should be and a broad consensus is starting to emerge from both Europe and the United States. There is now a need to move towards demonstrable implementation examples in practice.

The development of generalisable models of computable CPRs is of questionable value in isolation however. Many would say that the world already has more than enough ICT models! Additional value to this work can be added by also addressing more efficient forms of deriving the clinical knowledge itself that can then be used to dynamically populate those models. There is a need to move away from the existing manual forms of evidence population derived from literature-based sources with a view to demonstrating the feasibility of deriving valid clinical knowledge from electronic sources that form a basis for dynamic methods of knowledge translation. For reasons that will be discussed later, the goals of the emerging research around the LHS suggest that it is a good fit and provides a suitable context and platform within which an actual implementation can demonstrate the goals of this research in practice. As a starting hypothesis I suggest that the CPR development lifecycle can be considered to be a more specific instance of a broader LHS like infrastructure.

Three research questions are framed in this work around the two complimentary core functions of the LHS that together form an iterative learning cycle: to both *generate* and *disseminate* valid clinical knowledge into clinical practice. The first
Research question will address the requirements for development and implementation of computable models to represent eCPRs, in a format that promotes easier translation into primary care practice than their literature based equivalents. The second research question will address the need to derive valid clinical knowledge from electronic sources of patient data that can be used to populate and disseminate computable models of eCPRs. The third research question investigates the feasibility of consolidating representation and dissemination of CPRs as part of an electronic lifecycle supported by a learning health system infrastructure. This infrastructure can provide eCPR based evidence as a web based service, allowing third party tools to work effectively with electronic versions of CPRs to provide decision support. On that basis there can now be a formal statement of my three research questions:

**Research Question 1**—Can traditional literature based CPRs be represented in a computable, generalisable and interoperable format using open technical standards?

**Research Question 2**—Can clinically valid eCPR knowledge be dynamically developed using aggregated sources of primary care electronic health record data?

**Research Question 3**—Can a Learning Health System infrastructure support the derivation and dissemination of eCPRs as a service for use by third party tools providing diagnostic decision support?

To summarise this section and as will be demonstrated in the literature reviews to follow, the stated research questions have been derived on the basis of the following research observations:

- CPRs in their traditional forms have been shown to be effective but wider uptake has been hindered by the difficulties progressing them beyond initial derivation as part of the CPR development lifecycle
- CPRs are largely literature based and as such are not embedded within a defined clinical workflow or process
- current efforts at implementing specific eCPRs are not yet widely generalisable and interoperable
• generation of clinical evidence from electronic sources of patient data to support eCPR construction would represent a significant advance on the static CPR development lifecycle
• an electronic data driven eCPR development lifecycle can be shown to be a specific form of a more general LHS like infrastructure.

1.4 Aims/ Objectives of this Research
Based on the research questions a number of conceptual and implementation aims have been defined to drive this research. The conceptual aims are focussed on providing a suitable theoretical framework upon which a technical implementation can be demonstrated. The implementation aims describe critical implementation features to demonstrate a viable technology driven CPR development process. For clarity these aims are summarised in the following tables 1-1, 1-2 and 1-3. These tables clearly show the relationship between the research questions, identified gaps in knowledge and aims. Each aim has been allocated a number related to the research question number and will be used as a means to track their implementation progress in later sections.
Research Question 1 - Can traditional literature based CPRs be represented in a computable, generalisable and interoperable format using open technical standards?

Knowledge Gaps

- CPRs in their traditional forms have been shown to be effective but wider uptake has been hindered by the difficulties progressing them beyond initial derivation as part of the CPR development lifecycle
- CPRs are largely literature based and as such are not computable or embedded within a technology driven diagnostic process
- Current research efforts implementing individual electronic CPRs are not generalisable to other rules or demonstrably interoperable

<table>
<thead>
<tr>
<th>No.</th>
<th>Aim Description</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Develop a theoretical framework describing a dynamic eCPR development and dissemination process, consistent with the wider goals of the LHS that provides a vocabulary to discuss and develop a technology driven vision for eCPR development.</td>
<td>Conceptual</td>
</tr>
<tr>
<td>1.2</td>
<td>Describe more ambitious diagnostic strategies using decision support tools that address the limitations of traditional CPR dissemination to facilitate wider clinical acceptance of CPRs</td>
<td>Conceptual</td>
</tr>
<tr>
<td>1.3</td>
<td>Develop a generic model for the representation of computable eCPRs</td>
<td>Implementation</td>
</tr>
<tr>
<td>1.4</td>
<td>Develop an infrastructure based on openly available tools and technology standards based on proven implementations used in other commercial domains</td>
<td>Implementation</td>
</tr>
</tbody>
</table>
Research Question 2 - Can clinically valid eCPR knowledge be dynamically developed using aggregated sources of primary care electronic health record data?

Knowledge Gaps

- no existing research describes implementations demonstrating how generation of clinical evidence from electronic sources of patient data to support eCPR derivation might be achieved.

<table>
<thead>
<tr>
<th>No.</th>
<th>Aim Description</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Exploit existing repositories of electronic patient data for the purposes of demonstrating how a clinically valid diagnostic eCPR can be electronically derived and created in practice</td>
<td>Implementation</td>
</tr>
<tr>
<td>2.2</td>
<td>Validate the electronically derived eCPR against existing clinical literature to assess its agreement and consistency with current clinical knowledge</td>
<td>Implementation</td>
</tr>
</tbody>
</table>
Research Question 3 - Can a Learning Health System infrastructure support the derivation and dissemination of eCPRs as a service for use by third party tools providing diagnostic decision support?

Knowledge Gaps

- the manually intensive delivery mechanisms traditionally used to enable CPR development has not promoted successful dissemination in clinical practice; an LHS-like infrastructure is yet to be investigated as a suitable electronic platform to address these derivation and dissemination problems.

<table>
<thead>
<tr>
<th>No.</th>
<th>Aim Description</th>
<th>Aim Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Demonstrate the viability of service-based provision of electronic diagnostic evidence through the construction of prototype implementations of the theoretical eCPR framework</td>
<td>Implementation</td>
</tr>
<tr>
<td>3.2</td>
<td>Demonstrate the application of this technology to provide decision support using third party tools in practice</td>
<td>Implementation</td>
</tr>
</tbody>
</table>

1.5 Summary of Methods and Thesis Structure

The aims of this research as previously described are primarily practical software related outputs that demonstrate the proposed theory in practice. On that basis the primary research methodology employed is a functional prototyping approach, backed up by theory developed from extensive literature review.

The two primary outputs of this research are:
• the development of a theoretical framework to describe the necessary components of an infrastructure supporting derivation and dissemination of eCPRs
• the development of a fully functional prototype to demonstrate implementation of the theoretical framework in practice.

The following steps have been carried out and will be described in detail in the subsequent listed sections:

• Section 2: literature review to establish current state of the art regarding clinical applications of CPRs; as these topics are dynamic and relatively new emerging areas of technological research, the literature reviews have primarily focussed, though not exclusively, on more recent publications in the last ten years.
• Section 3: literature review to establish current state of the art regarding electronic applications of eCPRs to provide decision support
• Section 4: development of conceptual framework for implementation and discussion of eCPRs based on conclusions of literature reviews
• Section 5:
  o construction of clinical content for the specific scenario of Urinary Tract Infection, a condition that is common in primary care practice with extensive existing clinical evidence in the form of diagnostic guidelines available for direct comparison
  o application of conceptual framework in practice to produce a high-fidelity prototyped backend eCPR evidence service with appropriate toolsets
  o implementation and evaluation of data mined eCPRs based on comparison to literature
  o demonstration of applications of technology as applied by third party tools.
• Section 6: discussion and summary of research done.
2 State of the Art – CPR

Development in Primary Care

2.1 Objectives of Clinical Literature Review

The clinical literature review aims to develop a clear perspective on the clinical application of CPRs in practice. In order to develop suitable computable models and tools to represent diagnostic CPRs it is necessary to fully understand their current clinical applications and limitations. The purpose of the literature review therefore is to:

- understand the potential causes of diagnostic error in primary care
- define the role of CPRs in applying evidence-based medicine in practice to reduce diagnostic error
- fully define what a CPR is and the current traditional lifecycle for developing, disseminating and validating them in practice
- describe the broader role of CPRs within a defined diagnostic process in primary care practice with a view to enhancing patient safety through reduced diagnostic error.

2.2 What do we Mean by the Term ‘Primary Care’?

The efficient provision of clinical care is not delivered using a ‘one size fits all’ approach. Broadly speaking the frontline of clinical practice in Europe is supported by family practitioners working directly in the community who act as ‘gatekeepers’ directing patients to appropriate specialist care services found in the wider health system. I will refer to this gatekeeper role using the commonly used term ‘primary care’.

This research will focus exclusively on the primary care clinical setting. Secondary care or hospital services are not within the scope of this work. It is therefore necessary to define at the outset what exactly is meant when referring to the term ‘primary care’. A number of definitions exist which focus on provision of primary care from subtly different perspectives that focus on the process or the clinician concerned, and may interchangeably use the term ‘general
practice’ instead (35). For the purposes of this work the following two definitions from Starfield and the Institute of Medicine capture the important context.

Primary care is the

- ‘level of a health service system that provides entry into the system for all new needs and problems, provides person-focused (not disease-oriented) care over time, provides care for all but very uncommon or unusual conditions, and co-ordinates or integrates care provided elsewhere by others’ (36)
- ‘provision of integrated, accessible health care services by clinicians who are accountable for addressing a large majority of personal health care needs, developing a sustained partnership with patients, and practising in the context of family and community’ (37).

Based on these definitions we can see that primary care practitioners hold a unique and challenging position within the healthcare system. They are often the first point of contact for newly presenting clinical problems. They typically have a continuous relationship with patients and act in a co-ordinating role for access to other healthcare services.

Because of the ‘gatekeeper’ role that primary care practitioners provide, they must deal with a much wider variety of clinical presentations than their specialist equivalents in secondary care. This wide variation in clinical presentations combined with the need for good communication and integration with secondary care services presents a unique challenge for primary care practitioners and sometimes the process can fail (2).

2.3 Causes and Implications of Error in Primary Care

Adverse events in primary care practice occur for a wide variety of reasons. In order to contextualise adverse errors it is useful to clearly understand the practice of clinical reasoning itself as well as the primary care process in which it takes place. The cognitive act of clinical reasoning itself may contribute to causing adverse events (38, 39). On that basis I describe the ‘what, why and when’ of adverse events in primary care practice.
2.3.1 What is an ‘Adverse Event’?

Incidents that occur as part of day-to-day clinical treatment that ultimately result in a negative impact on the clinical outcome of any particular patient are commonly referred to as ‘adverse events’ (2). It is necessary to identify the most problematic areas of patient treatment that could potentially result in an adverse event taking place in order to effectively propose solutions that contribute positively to patient safety. Where an adverse event occurs there is an associated impact on the patient resulting in a decrease from optimum levels of patient safety to a varying degree depending on the seriousness of that adverse event. The following definition provides a working description that interestingly also includes incidents which may have the potential for harm to a patient to take place:

- ‘An unintended event, no matter how seemingly trivial or commonplace that could have harmed, or did harm a patient’ (40).

Previous systematic reviews have focussed on analysis of clinician reported adverse events in the primary care context and have proposed useful classifications of the major underlying problem areas that have been identified (2). Classifications proposed by Elder and Dovey and others attempt to address the two key questions of where adverse events occur most often in primary care and why those adverse events may have occurred (2, 41).

Elder and Dovey identified and classified three main problem areas in relation to the question of where adverse events take place within the primary care process. These are:

- diagnosis – errors relating to misdiagnosis through either missing a diagnosis entirely or mistakes which result in a delay in reaching a final correct diagnosis
- treatment – errors relating to incorrect provision of treatment to a patient broadly broken into two areas of drug and non-drug treatment
- preventive services – errors relating to incorrect provision of preventive services to patients.
A fourth catch-all category of ‘other’ was proposed to cover what might be considered ‘system’ non-clinical causes of error in the original landmark discussion of error in the health system (41). These errors may relate to organisational factors that are not directly under the remit of clinical judgement e.g. loss of a radiology document related to a patient resulting in delays to treatment.

The classification of errors as proposed by Elder and Dovey has also been consistent with other studies that have attempted to quantify the relative contribution of each error category. A study of 805 incident reports submitted by GPs in Australian primary care practice found 51% related to pharmacy treatments, 42% related to non-pharmacy treatments, and 31% related to diagnostic decision making (40). A review of malpractice claims found diagnostic error to be the most reported reason in 23-63% of claims, with pharmacy treatment errors cited in 5.6-20% of claims (42).

With a clearer understanding of the potential causes of adverse events we can begin to develop solutions which aim to improve patient safety in the context of provision of primary care. As outlined previously, the first step is to identify the major causative areas of adverse events that are found to occur in primary care. The second step is to then analyse these in more detail to establish why adverse events may be happening within each of these major causative areas.

2.3.2 Why do Adverse Events Happen?
In relation to the question of why adverse events take place within the clinical process, four potential factors for further investigation are described by Elder and Dovey. These are:

- clinician factors – factors directly attributable to decisions made based on the judgement and evaluation of the clinician regarding a particular case
- communication factors – factors relating to communication failures or misunderstanding between clinical entities collaborating as part of the overall patient treatment process
• administration factors – factors attributable to administrative errors in patient case management either by the clinician, other associated staff or third parties collaborating with them

• external factors – impacting factors not directly in the control of the clinical consultation itself e.g. regulations, social and family factors.

This classification as proposed is useful but, as stated by the authors, does not attempt to quantify the individual contribution of each of these factors to negatively impact on patient safety. Other studies have taken the logical next step and focussed on a specific identified category of error, such as clinician diagnostic error, and attempted to assess the contribution of that specific category of error as potential threats to patient safety (3).

In the Kostopoulou study, adverse events in the form of diagnostic errors were identified as having the potential to result in the most serious consequences for the patient of all the previously identified error categories (2). Further investigation looked at the properties of diagnoses that were characterised as being potentially problematic for clinicians in forming a correct final diagnosis. It is logical to develop our understanding about why these diagnoses present in a more difficult fashion in the context of the diagnostic process as traditionally employed in a clinical consultation. These were found to be (3):

• atypical presentation – presenting features of illness are different to classic features normally associated with that illness

• non-specific presentation – no presenting features unique in distinguishing the illness resulting in similar presentation to many other differential diagnoses

• very low prevalence – uncommon illness in the general population where primary care clinician is not used to seeing characteristics of such a rare illness on a regular basis (the EU definition of a low prevalence disease is a ‘prevalence of less than 5 per 10,000 in the community’) (43)

• co-morbidity – diagnoses that may be accompanied with other complicating illness
perceptual features – illness where diagnosis is strongly indicated by visual or auditory cues rather than measureable, testable or symptomatic features.

I therefore conclude that adverse events related to the primary care clinician of a diagnostic nature may be more likely to take place where more unusual clinical cases present that do not have classic associated features that easily distinguish the illness in question. In the case of rare conditions the knowledge base available to any particular clinician may be limited by their own case experience. This may impact on the initial formulation of a suitable working diagnostic hypothesis which has been shown to be a crucial first step in the diagnostic process (44).

These difficulties can be further compounded by the commonly held belief that the more experience a clinician has gained through frontline clinical practice, then the better the quality of decision making. Clinical pattern recognition based on previous clinical case history has long been recognised as valid and useful in identifying a potential diagnosis based on the presenting characteristics of a patient. It has been argued however that there may be an inverse relationship between length of clinician experience and quality of provided (45). Diagnosis based on previous case history may not be as successful where case presentations occur that are outside the more commonly presented cases for any particular clinical environment. In these situations the potential for a diagnostic adverse event may increase.

The area of diagnostic error has serious consequences for patients and clinicians because it is more difficult to define and quantify than other categories of error. Because of this ambiguity it is also an area that frequently gets overlooked when developing patient safety and quality strategies (46). The challenge therefore is to deal with the wide variation in the unique characteristics of each particular clinical case that can contribute to their diagnostic difficulty. Difficulties arise where cases may fall outside of the experience of the cognitive knowledge base used by the clinician. Ultimately this can manifest itself in damage to patients, healthcare professionals and the healthcare system itself. Studies in both the US and Europe have shown that
diagnostic error is the most common cause for litigation in primary care and as such also represents a serious financial burden on the health system as well as a serious patient safety issue that needs special attention (47-49).

2.3.3 When do Adverse Events Happen?

I have now highlighted diagnostic error as a threat to patient safety in primary care and identified the characteristics of certain diagnoses that are particularly problematic. How can tools be developed that can assist primary care clinicians with the diagnostic process so that they might consider diagnoses that might otherwise potentially be missed where they present in rare or difficult cases? In order to propose potential solutions to address these issues it is necessary to understand the diagnostic process in primary care itself as well as the steps and decision points that take place within it. Only then can we begin to suggest possible solutions to address the points of failure highlighted.

Diagnostic error can occur at a number of distinct points in the overall diagnostic process (38). These errors can be system related failures as part of the operation and conduct of the health system itself, or cognitive failures that manifest themselves as a result of clinician judgement. Cognitive failures can occur during the:

- gathering of patient data relevant to the case (diagnostic workup)
- reference of existing clinician knowledge pertinent to a case (the knowledge base)
- synthesis and application of existing clinician knowledge with the particular patient case data (diagnostic reasoning) (50).

Flawed clinician judgment has the potential to result in either a delayed or missed patient diagnosis that results in a poorer patient prognosis. This can occur on the basis of either presenting symptoms relating to an existing condition or opportunities for prevention of the future occurrence of a condition.

Previous studies have deconstructed the diagnostic process with a view to understanding the sequence of steps that take place in order to arrive at a final confirmed diagnosis (51). These are useful with a view to identifying at what
point potential electronic solutions may best be deployed to provide assistance as part of the diagnostic process.

In a study by Heneghan et al. the following sequence of diagnostic steps was identified and is shown in figure 2-1:

- initiation of investigation of diagnostic hypotheses
- formulate diagnostic hypotheses in the form of potential differential diagnoses
- refinement through rule in or rule out hypotheses by history and examination
- define and select final diagnostic hypothesis
- confirm final diagnosis through test of time, trial treatment or a definitive gold standard test (51).

![Figure 2-1 Diagnostic stages and strategies employed in primary care – courtesy of Heneghan et al.2009 (51)](image)

As part of each step a number of diagnostic strategies are employed. The Hengehan study asked general practitioners to highlight the strategies they use in actual practice to make an initial diagnosis (figure 2-2) and to subsequently refine the diagnosis (figure 2-3). Unsurprisingly the most important consideration at the outset of formulation of a diagnostic hypothesis is the actual presenting patient complaint. An alternative description for this is the
‘Reason for Encounter’ (RFE) which I will use to refer to this from now on in this work.

Once a set of potential diagnoses for consideration have been suggested a process follows of refinement. The two most widely reported strategies applied were the use of ‘pattern recognition’ and ‘restricted rule outs’. Pattern recognition typically relies on identifying a similarity between the pattern of diagnostic cues shown by any individual presenting case, and those recalled as significant in previously confirmed diagnoses of other patients. This technique relies heavily therefore on clinician case history experience (50).

Restricted rule outs employ a strategy of refining differentials to consider based on identifying diagnostic cues that can provisionally rule out the possibility of a differential unless subsequent consultation information is found that could rule it back in. For example the absence of lower right quadrant pain would provisionally rule out a diagnosis of appendicitis. This technique is based on a process of elimination to use rule outs to eliminate unsuitable diagnoses with a view to leaving a manageable few diagnoses that can be followed-up, with clinical testing if required, in greater detail (44). The least used strategy was the application and use of CPRs which will be discussed in detail later.

![Figure 2-2 Strategies used by general practitioners in making an initial diagnosis - courtesy of Heneghan et al. 2009 (51)](image-url)
2.3.4 When Diagnosis in Primary Care Goes Wrong

The successful application of these recognised diagnostic strategies is not guaranteed and the literature regarding malpractice claims in family practice would suggest that there are significant problems to be addressed in this important area. The systematic review by Wallace et al. highlighted diagnostic error or delay as representing 23%-63% of claims depending on the study (42).

The financial burden and cost of diagnostic error in Ireland is difficult to assess as adverse event reporting and a more general patient safety culture are not well developed and coordinated in our health service as highlighted by the recommendations of a recent Health Information Quality Authority report on patient safety surveillance and intelligence (52). There has been some limited claims data relating to hospital adverse events as reported to the State Claims Agency. This report suggests that adverse event reporting is under-reported in Ireland with an adverse event rate reported of 2.9% compared to 4-16% internationally (53). Of the 76,842 claims reported to the State Claims Agency for 2012, only 1,364 (1.8%) of these were reported as diagnostic errors.

An analysis of malpractice claims in the US however was consistent with the higher end of the international figures and also gives us an indication of where
the problems arise in primary care diagnosis and the financial implications of this (47). In the Controlled Risk Insurance Company (CRICO) analysis of 2,596 malpractice claims recorded between 2008 and 2012, it was found that 58% of those claims were directly attributable to errors made in the initial diagnostic assessment, a figure that is similar to the higher figures found in the UK (48). As shown in figure 2-4, this was broken down further to show that of those claims, 33% of those cases were attributable to problems related to the initial diagnostic workup, and 34% were attributable to errors made in the process of formulating a suitable set of differential diagnoses to consider. The remaining cases were attributed to problems related to clinical testing and subsequent patient follow.

![Diagram](image)

*Figure 2-4 CRICO analysis of where errors were made in diagnostic assessment of 2,596 malpractice claims, CRICO 2015 (47)*

The financial implications of diagnostic error as described in the CRICO report are truly immense, particularly when considered in the current climate of ‘austerity’ that has imposed tighter controls on healthcare spending. Of the 2,596 malpractice cases reviewed, 909 cases were closed with a financial payment of compensation. The average cost per case to settle was $442,000 (47). This equates to a total cost of €401,778,000 over four years, giving an
average yearly cost of $100,444,500 for just these analysed cases in the US. These are costs that are indirectly incurred by all patients in the health system in the form of rising health insurance premiums; costs that are directly attributable to errors made at the very start of the patient clinical journey when carrying out the initial diagnostic assessment in primary care. These costs represent a huge financial burden at a time when health systems can least afford it.

The formulation of all potential differential diagnoses to consider at the outset based on the underlying presenting problem is therefore crucial in ultimately making the correct final diagnosis. Equally important is the process of refinement based on pattern recognition and eliciting associated cues that supports a process of ruling out inappropriate diagnoses. These processes and the reasons for how they might fail are primarily cognitive tasks. I will now consider how cognitive bias can distort these strategies resulting in diagnostic error.

2.4 Cognitive Strategies Used for Diagnostic Decision Making

Cognitive strategies for diagnostic reasoning can be categorised according to the more widely applicable models of ‘slow’ and ‘fast’ thinking (alternatively referred to as ‘analytical’ or ‘non-analytical’ thinking) (54-56). The choice of whether to use ‘slow’ or ‘fast’ diagnostic strategies is influenced by the perceived difficulty of the presenting case, and also by the clinical experience of the diagnosing clinician (50).

There is evidence that less experienced clinicians will practice more defensively and rely more on ‘analytical’ diagnostic testing to begin with. As the experience of the clinician grows in the form of a larger body of previous case histories to draw on, ‘non-analytical’ strategies become used more often (44). Primary care practitioners also have to work within limited consultation times to reach their conclusions so the practice of analytical methods in every presenting case may also not be practical.
2.4.1 Intuition and Heuristics as Part of the Diagnostic Process

I have previously shown that clinicians use a number of different diagnostic strategies when conducting consultations with their patients (51). One example discussed was the use of ‘pattern recognition’ where experienced clinicians may draw on a successful diagnosis confirmed in a previous case history to spot a newly presenting case in a different patient. This is an example of a diagnostic strategy that relies on intuition or what is more formally referred to as heuristics.

The use of intuition and heuristics by clinicians as part of diagnostic reasoning is an accepted if not formally understood diagnostic strategy (57, 58). It has not always sat comfortably beside the concept of practising evidence based medicine but is certainly not incompatible with it (59). The complexity of diagnostic reasoning has meant that clinicians employ shortcuts or ‘rules of thumb’ in order to make the initiation of the process quicker using a ‘top-down’ approach to diagnosis that employs these shortcuts. These approaches may not be easy to formally define. A number of mechanisms may be at work that involves application of:

- pattern recognition where previous case experiences are intuitively used to identify potential matching diagnoses to consider
- ‘gut feeling’ where the presenting complaint (reason for encounter) of the patient intuitively informs an avenue of diagnostic investigation (57).

The difficulties of exclusively pursuing alternatives to heuristic approaches are articulated by Cook where he states that ‘arguably, without a working knowledge of gestalt principles, clinicians would be hopelessly bogged down with “bottom up” assessments of their patients, begrudgingly ploughing through reams of clinical data to form a workable hypothesis’ (60).

2.4.2 Cognitive Bias in Heuristics

The reliance of heuristics on a uniquely defined clinical variable in the form of clinician ‘case experience’ allows us to conclude that it cannot be free from problems of consistency (61). The application of heuristics by different clinicians in relation to the same presenting patient case can only deliver identical conclusions, and will not result in an incorrect avenue of diagnostic investigation
to begin with, in unlikely circumstances; when an identical reasoning process has taken place using identical evidence bases available to each individual clinician as defined by their case experience and clinical education. The likelihood of consistency of reasoning in the same clinical cases reduces further when we consider other factors collectively known as 'cognitive bias' that have potential to lead to 'information distortion' in the diagnostic process (61). For the same presenting patient therefore we can state that two different clinicians may pursue different clinical case workups.

I have previously discussed the sequence of iterative steps that takes place as part of the clinical consultation:

- step 1 - formulation of differential diagnoses to consider based on the presenting patient problem
- step 2 - refinement of differentials through rule out unlikely diagnoses and framing a leading diagnosis to consider
- step 3 - detailed follow-up and investigation of a leading diagnosis (44, 51).

Each of these steps may be subject to different types of cognitive bias (57, 58, 62). Once a leading diagnosis has been selected, information distortion can take place as a drip-feed of diagnostic cues are fed into the diagnostic process. This may result in the oversight of pertinent clinical evidence that is subsequently found to contradict the leading diagnostic hypothesis in order to accommodate the initial false leading diagnosis (62). In the specific case of diagnostic reasoning cognitive bias may arise in each of the diagnostic steps as follows:

- ‘anchoring bias’ in step 1 or step 2 – the tendency to frame a leading diagnosis to consider by placing undue importance on the first piece of diagnostic information presented by the patient
- ‘availability bias’ in step 1 or step 2 - choosing a leading diagnosis based on previous case experience or recently read clinical research related to a ‘memorable’ case that might fit the newly presenting case pattern whilst giving undue consideration to other possibilities
• ‘confirmation bias’ in step 2 – the consideration of new diagnostic evidence giving undue importance to only the evidence that ‘fits’ the already formulated leading diagnosis, whilst giving undue consideration to evidence that may contradict it

• ‘premature closure’ in step 3 – the formulation and decision to conduct detailed investigation of a leading diagnosis that first fits the presenting case without giving due to consideration to other possible differentials that also fit the case (54).

It needs to be acknowledged that, based on estimated diagnostic error rates in primary care of 10-15% of consultations (40), Heuristic methods still have value in clinical practice and a combination of applying non-analytical and analytical methods has been advocated by some (63, 64). There is still room for improvement and the development of analytical methods of clinical practice such as evidence-based medicine can however be considered a direct response to address the limitations inherent in heuristic clinical practice.

2.5 Improving Diagnostic Decision Making through Practice of Evidence Based Medicine

Relying solely on previous experience based on pattern recognition of case histories may not be sufficient for diagnoses with problematic presenting characteristics discussed previously. In these cases a rigorous process of evidence review needs to take place that is based on the latest available peer-reviewed clinical evidence. Such a process informing the consideration of differential diagnoses to consider at the outset may ensure that a correct final diagnosis is not ultimately missed resulting in harm to the patient.

From a technology perspective, the goal is not to attempt to replace clinician judgement but to support rigorous evidence evaluation as part of the diagnostic process. This process can be considered to require development of tools that assist with the formulation of appropriate differential diagnoses through effective practice of evidence-based medicine.

Evidence-based medicine has been advocated as one way of helping to ensure that diagnostic reasoning is based on a more rigorous and systematic
approach (8, 22). It has also been argued that the value of evidence based medicine should not be overstated and incorrectly presented as a ‘silver bullet’ for addressing all of these issues on its own (6, 59, 65-67).

2.5.1 Evidence Based Medicine Evolves

The most widely cited definition of evidence-based medicine was proposed by Sackett and describes it as ‘the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients’ (22). This definition hinges on how you might define and quantify the term ‘best evidence’. This is a challenge that is also being dealt with in other public service domains. The controversies of quantifying the value of education and teacher’s performance are eloquently expressed by the headmaster regarding one of his teachers in Alan Bennett’s History Boys. This could just as easily express the frustrations articulated by some primary care practitioners regarding the motives of the evidence-based medicine movement:

‘It isn’t that he doesn’t produce results. He does. But they are unpredictable and unquantifiable and in the current educational climate that is no use. He may well be doing his job, but there is no method that I know of that enables me to assess the job that he is doing’ (68).

The definition of evidence-based medicine has therefore evolved to specifically address the requirement to empirically quantify what constitutes ‘best evidence’ (69, 70). More recent definitions have focussed on being more explicit about the role and use of mathematical models to quantify risk, benefit and harm when making a clinical judgement: ‘evidence-based medicine is the use of mathematical estimates of the risk of benefit and harm, derived from high-quality research on population samples, to inform clinical decision-making in the diagnosis, investigation or management of individual patients’ (8). It is therefore useful to explicitly state that it is this more recent interpretation and definition of evidence based medicine that we use in the context of this research. As I shall discuss in detail next, the development of CPRs derived from population samples in the form of observational cohort studies is one such tool that is consistent with this newer definition of evidence-based medicine in practice.
2.6 What are CPRs and How are they Applied?

It is necessary at the outset to clearly define what we mean when we refer to a CPR. A CPR ‘is a clinical tool that quantifies the individual contributions that various components of the history, physical examination, and basic laboratory results make toward the diagnosis, prognosis, or likely response to treatment in a patient’ (23, 71).

2.6.1 Types of CPR

As has been stated in the definition, although all CPRs share a common structure they may have different clinical applications. CPRs may be related to diagnostic, prognostic or treatment based outcomes. As such, from a strictly diagnostic perspective a CPR quantifies the contribution of what I will refer to from now on as ‘diagnostic cues’ to a diagnostic outcome. These ‘diagnostic cues’ are in the form, as originally stated for the general definition, of ‘components of the history, physical examination, and basic laboratory results’. These include patient reported diagnostic symptoms, such as patient reported pain, along with clinician observed diagnostic signs, such as fever confirmed by a clinician measured temperature.

2.6.2 CPRs as a Form of Probabilistic Diagnostic Reasoning

We have previously seen that approaches to cognitive reasoning can be broadly classified as non-analytic or analytic. Diagnostic CPRs constitute an analytic approach to diagnosis based on probabilistic diagnostic reasoning (also referred to as Bayesian reasoning based on Bayes Theorem) (70). The required components to apply probabilistic diagnostic reasoning when trying to establish a positive diagnosis for a specific patient presenting with a diagnostic condition under investigation are (72):

- pre-test probability - the current prevalence of the target condition as found in the presenting patient population
- newly presenting diagnostic evidence - diagnostic cues elicited from the presenting patient that are known to quantifiably increase or decrease the probability of a positive diagnosis related to the target condition
post-test probability - the adjusted probability of a positive diagnosis related to the target condition, once consideration has been given to new quantifiable diagnostic evidence applied to the original pre-test probability.

On consideration of all the presenting patient evidence, in the form of individual diagnostic cues with a quantified contribution to a positive target diagnosis, the final post-test probability may be significantly greater or lower than the original pre-test probability of the target condition under investigation. These probabilities will always be between 0 and 1. A post-test probability of 0 represents certainty that a positive diagnosis of the target condition can be ruled out and 1 represents certainty that a positive diagnosis of target condition is established and suitable treatment should commence (72).

2.6.3 CPRs as a Form of Threshold Based Risk Interpretation and Action

A less certain and more realistic situation is where post-test probabilities are found in any of the range of possible values between 0 and 1. In this case we need to interpret the clinical significance or risk along with the appropriate clinical response to any such observed value.

The application of what is known as a ‘threshold’ approach to clinical risk interpretation and subsequent action has been advocated as a more clinically useful alternative to focussing on the absolute ruling in or ruling out of a diagnosis on its own (73). This approach defines risk bands that span the full range of possible post-test probability values between 0 and 1. Each band defines a non-overlapping upper and lower threshold to which a clinical or risk interpretation is attached (72).

From a diagnostic perspective, the appropriate setting of these thresholds, which determine the probabilistic level of certainty required to confirm a positive diagnosis or rule out a negative diagnosis, is determined by clinical factors associated with subsequent treatment or testing including:

- prognosis of the diagnosis
- availability of appropriate treatment/testing for the diagnosis
- safety of subsequent treatment/testing for the diagnosis
• costs of subsequent treatment/testing for the diagnosis (72).

This allows the definition of a manageable scale of thresholds in the form of discreet ranges of possible post-test values. Rather than placing undue significance on each and every possible post-test probability value in itself, risk categories can be associated with each defined risk threshold along with a recommended clinical action. The combination of associated thresholds, risk categories and actions provide for manageable evaluation of the significance and clinical interpretation of any individual post-test probability result (69, 72).

2.6.4 The Formal Structure of a CPR

The formal characteristics of CPR can be clearly identified based on the previous CPR definition and the associated discussion on probabilistic reasoning. The quantification of the contribution of diagnostic cues to a diagnostic outcome comes in the form of a derived scoring scheme for each CPR. The scoring scheme also defines threshold based score ranges with associated decision outcomes that are related to diagnosis or prognosis (23).

A CPR is derived from a statistical model and will be constructed and structured based on the following distinct parts (70):

• a clinical outcome that relates to a defined diagnostic, prognostic or treatment outcome associated with a selected clinical condition
• a set of diagnostic cues and associated criteria that are quantifiably clinically significant to the clinical outcome being assessed by the rule
• a statistically derived scoring scheme that quantifies the relative contribution of each cue where present or absent in relation to the clinical outcome
• a threshold based scoring scheme that defines relative clinical interpretations of risk categories for all possible scores for the rule
• an optional decision indicating a clinical action in response to each risk category to be recommended based on each of the defined threshold scores.
It is clear therefore based on this definition that the practice of evidence-based medicine that makes use of underlying mathematical models includes the use of CPRs in clinical practice.

2.6.5 The Derivation Process of a Diagnostic CPR

CPRs are typically derived through the conduct of manually intensive observational studies in the form of cohort or cross-sectional studies (74). From a diagnostic point of view, these studies are conducted to investigate and identify potential epidemiological associations between suspected diagnostic variables and a defined diagnostic outcome. These associations may then be quantified statistically to establish a measurement of the degree of correlation (the measure of association of linearly related variables) with a view to identifying candidate associated variables that might be used as predictors of the diagnostic outcome (8).

A subset of strongly correlated diagnostic input variables may subsequently be chosen to construct a predictive model to establish the post-test probability of a positive diagnostic outcome based on a set of presenting patient diagnostic cues. The goal is to construct a parsimonious model that maximises predictive power whilst reducing as much as possible the total number of input model input variables (75).

Many predictive model types may be used but typically for CPR construction they are in the form of a logistic regression model. A regression model defines a mathematical equation to predict a linear directed relationship between one or many target and dependent variables. These models predict a diagnostic outcome based on the input of a single diagnostic variable, or more likely the input will be more than one diagnostic variable (known as a multi-variable logistic regression model). The construction of such models as part of research studies should be subject to best practice guidelines such as the ‘Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis’ (TRIPOD) guidelines (76). These guidelines make explicit how the model was developed in terms of patient recruitment practice, analysis and refinement applied, along with the limitations and assumptions underlying the model itself. The output of the regression model is a set of odds ratios or
likelihood ratios that quantify the contribution of each diagnostic variable to the diagnostic clinical outcome. These quantified diagnostic variables can be rounded to make a user-friendly ‘score’ that can be used to assess any presenting case (70).

Finally a set of risk categories or threshold score bands are set. These risk categories define what clinical actions should take place based on any score within defined ranges. For example these may be defined in terms of risk interpretations such as ‘low’, ‘medium’ or ‘high’ where a low score means the diagnosis can be effectively ‘ruled out’, a ‘medium’ score is unclear and requires ‘further testing’ and a high score means high certainty of diagnosis and suggests to ‘commence diagnostic treatment’ immediately (70).

In summary the sequence of steps to derive and construct a CPR is:

- observation of suspected diagnostic variables of influence through conduct of a cohort or cross-sectional study
- establish quantified associations as correlations of diagnostic variables and identify ‘strongly’ correlated variables
- construct a predictive model on strongly correlated variables usually, though not exclusively, involving use of logistic regression modelling (70).
- establish thresholds diagnostic model in the form of score based rule to support probabilistic reasoning (70, 74).

It should be stated that the identification and selection of suitable input variables to construct a predictive model is a crucial initial step. Ideally the predictor variables selected should not be correlated as strong correlation results in collinearity. In this situation the overall outcome prediction may still be reliable but model results for the individual predictor variables may not be reliable when considered in isolation (75). In practice collinearity is present to some degree in most models and can be measured by the correlation coefficients among predictor variables (the ‘r’ value). If ‘r’ is greater than 0.8 for two predictor variables this may result in collinearity in the model. In such cases it may be appropriate to combine correlated variables into a single variable (75).
Once derived, a CPR then needs to be deployed and tested in practice as part of a defined CPR development lifecycle to see if the predictive model used is effective and reliable in wider clinical practice. The performance of a clinical prediction rule is traditionally assessed using two main characteristics, the calibration and discrimination of the rule (77). The calibration of the rule refers to how accurately the rule predicts risk in practice. From a diagnostic point of view the discrimination of the rule is more important since this establishes how well the rule predicts and differentiates the risk between patients with or without the clinical outcome under investigation (70). If the predictive value is more widely established as valid, a research trial may be carried out to establish the actual impact and benefit of using the rule in practice to change clinical practice behaviours (74).

2.6.6 The CPR Development Lifecycle

The development of a CPR typically follows a defined lifecycle, shown in figure 2-5 that indicates the maturity and wider applicability of the rule to more general patient populations (70, 74). The stages of development typically include:

- derivation – the initial construction of the CPR based on a statistical analysis of data from a defined patient population
- narrow validation – testing of the CPR against a similar patient population and clinical setting to the original derivation patient population to assess its performance in practice
- broad validation – testing of the CPR to a different clinical setting and patient population outside of the original derivation population to assess its validity and performance in a wider clinical context
- impact analysis – conduct of an RCT to assess the wider clinical impact of prolonged use of the CPR in clinical practice with a view to improving defined clinical outcomes or behaviours.
The illustration of the CPR development lifecycle in practice is best demonstrated using a clinical example.

2.6.7 The Ottawa Ankle Rule as a CPR Development Lifecycle Example

The application of the CPR development lifecycle in practice is discussed using the well known clinical example of a stage 4 mature CPR that has gone through the derivation, validation and impact analysis: the ‘Ottawa ankle rule’ is one such famous example (78).

The original clinical problem being addressed was to identify clinical features that were highly predictive of the presence of an ankle fracture to an extent that warranted subsequent investigation in the form of an x-ray. Prior to development of this rule, it was observed that x-rays on suspected ankle trauma patients in emergency departments were unnecessarily being carried out in many soft-tissue damaged cases and were subsequently found to be normal. The desired clinical impact of implementing such a rule was therefore to promote better use of radiography resources by reducing the number of unnecessary x-rays being carried out.

A derivation cohort study was carried out to identify predictive clinical features for ankle or foot fractures based on an analysis of 750 patients presenting with ankle trauma at an emergency department. The original derivation of the rule identified 17 significantly correlated features that predicted ankle fracture and warranted further x-ray investigation. A number of combinations of the significant predictor variables were tested in the form of predictive models before a final rule with 3 predictors that had 100% sensitivity and 40.1%
specificity in correctly predicting or ruling out ankle fractures for the original rule cases:

- age 55 or greater
- unable to bear weight on foot for four consecutive steps
- bone tenderness at posterior edges of either of the bony projections found on either side of the ankle (malleolus of ankle) (78).

In the retrospective hospital validation done on 689 suspected ankle fractures that were radiographed, the rule successfully identified all 70 cases of ankle fractures and would have resulted in a 36% reduction in unnecessary radiology requests (78).

A variant of these was also identified for identifying foot fractures warranting x-ray investigation. A validation of this constructed rule was carried out on a further 1032 patients that resulted in refinement of the initial rule (79). This refinement included the dropping of the age restriction as part of the original rule. Where a patient presented with the full set of diagnostic cues indicated by the rule, the risk of ankle fracture was high.

The subsequent clinical action was to follow up and confirm with an x-ray. The validation of the rule was successful resulting in a number of implementations of the rule in clinical practice by the original rule developers (80, 81). The impact analysis of implementing these rules found a significant decrease in the number of unnecessary ankle/foot radiography with associated cost savings. The success of these original studies resulted in a number of follow-on studies applying the rule in other clinical settings demonstrating broader validation and successful impact (82-85).

2.6.8 Barriers to Successful CPR Development

The success of the implementation of the Ottawa ankle rule and its subsequent broad acceptance as a working clinical tool in practice can be considered to be an exceptional example of a successful and maturely developed CPR. Not all CPRs are developed to this standard however and a number of barriers to successful CPR development have been identified.
Despite the existence of an accepted development lifecycle for producing CPRs, many studies describing them have traditionally focussed solely on the derivation phase of the CPR lifecycle (86). Others are subject to poor or non-existent CPR validation and impact analysis (34). Not all CPRs are therefore considered the same in terms of assessing their clinical acceptability and the relative strength of evidence and utility that can be attached to them.

CPRs that have only been subject to derivation or narrow validation are not yet considered suitable for use in wider clinical practice (71). This hierarchy of evidence broadly corresponds to the four steps of the CPR development lifecycle. Those CPRs that have been subject to all four steps including impact analysis are considered the strongest form of CPR evidence (71). Such CPRs can be used widely in clinical practice with a high degree of confidence in their accuracy and ability to positively change defined clinical outcomes.

Barriers to progressing lifecycle development of CPRs include (34, 71, 87):

- lack of validation of CPR and acceptability to clinicians
- lack of a suitable delivery mechanism to deploy an integrated CPR at point of care.

2.6.8.1 Lack of CPR Validation and Acceptability to Clinicians

Although many diverse examples of CPRs in primary care can be identified in research literature, their use has yet to gain widespread acceptance among clinicians (87). There are a number of valid concerns that influence why clinicians are reluctant to use them as part of their day-to-day clinical practice and sometimes they may simply be seen as a challenge to the clinicians’ autonomy to inform their own decisions (87).

Another more fundamental issue is clinical acceptability of the CPR. A survey of GP attitudes to using CPRs highlighted that their use of CPRs was mainly restricted to those that have been accepted and integrated into wider clinical guidelines (88). A CPR is unlikely to be accepted and integrated into wider clinical guidelines unless it has been developed according to accepted development standards and developed to at least broad validation (71).
Many derived CPRs are not fully compliant with the accepted methodological standards that have been identified as necessary to support good quality CPR derivation (89). Where those methodological standards have not been demonstrably followed or transparently reported, this raises questions relating to the validity of predictors used, the quality of the evidence and the confidence that clinicians can attach to the subsequent CPR recommendations made.

Even where GPs are aware of CPRs existing outside the context of approved clinical guidelines, the lack of validation of CPRs severely limits their perceived applicability to the same restricted patient populations defined in the original derivation research populations. As previously discussed a CPR that has not undergone any validation is not considered safe or usable in wider clinical practice. Even where narrow validation has occurred, the predictors that have been identified and the relative scores assigned to them for any derived CPR may be correct with respect to the original derivation and validation population, but may not be more widely applicable to the broader population (71).

Clinicians may also find that a CPR subject to derivation or narrow validation only may define predictors that are not clinically sensible with respect to the context of application in their own population or clinical setting. A CPR derived from a hospital population for example may define predictors based on testing that may not be timely or available in a primary care setting.

Complications may therefore arise when there are multiple rules derived by different researchers for any chosen clinical condition. As an example, a clinical condition such as Pulmonary Embolism has numerous variations of CPRs that may pertain to it (90). This can lead to confusion and a lack of clarity about which CPR variations are the ‘correct’ or ‘best’ ones to use, and two areas of difficulty in selecting CPRs may arise.

There may be more than one CPR available that is suitable to assess the risk of a given condition in a particular patient and clinical setting. These CPRs may have been derived from similar population samples and clinical settings as the patient and as such are valid tools to use. These CPRs however may have been defined differently with different scoring schemes or cues to be checked and as such may perform with different calibration or discrimination.
characteristics, despite being suitable choices for the patient population and clinical setting in question. The performance characteristics of one CPR in this situation may be significantly better than another despite both being suitable choices.

A more obvious difficulty arises where more than one CPR is available to assess the risk of a given condition in a patient but those CPRs have been derived from different clinical settings or patient populations and broad validation has not been successfully established. In this situation it may simply be incorrect to apply a given CPR derived from a given clinical setting, such as primary care, and apply it to a patient in another clinical setting such as secondary care where prevalence of the condition may vary and rule performance will be poor.

2.6.8.2 Lack of an Integrated Delivery Mechanism

The traditional delivery mechanism for CPRs has largely been though research literature, putting an onus on clinicians to search literature for suitable CPRs (33). This is compounded by the fact that literature based rules are by their nature static in content and do not provide for recording of versioned rule changes. This may have implications for the applicability of any particular CPR as changes take place over time in the demographics of the original rule derivation study population.

Reviews have also highlighted the importance of capturing the demographic context of the study population that has been used to derive a CPR. Score performance may vary when the CPR is applied to populations with different gender, age or clinical settings to the original derivation population. The Alvarado score for example has been found to perform best in adult male populations (91).

These changes in demographics may change the pre-test probability of the diagnostic condition along with the discrimination and calibration of the rule. This implies that a more dynamic approach to CPR derivation is needed that allows them to ‘learn’ and be revised as their performance characteristics change with the underlying derivation population.
The current highly manual nature of deriving CPRs makes them difficult to use and maintain (25, 33). Impact analysis of CPRs is typically established using a randomised control trial to establish the impact of the integrated CPR as an intervention in practice (12, 71). An essential part of carrying out such a trial is the identification of a suitable workflow integrated delivery mechanism for CPRs that can form the basis of the required study intervention (74). The integration of CPRs as part of clinician workflow potentially involves multi-disciplinary coordination and agreement among a number of different organisational groups where the CPR is to be deployed in clinical practice (26).

One increasingly researched way of delivering workflow integrated CPRs at point of care is through development of clinical decision support systems (CDSSs) based on computable models of clinical evidence (12, 15, 25, 92, 93). The barriers and facilitators in using ICT as an intervention delivery mechanism will be discussed in detail later in sections 3.2.4 and 3.2.5 as part of the ICT literature review.

2.6.9 **Application of CPRs as Part of a Diagnostic Strategy**

The process of making a diagnosis is simply one clinical decision point that needs to take place in a broader sequence of diagnostic events. Musen emphasises this point and describes the diagnostic process as the act of ‘deciding which questions to ask, tests to order, or procedures to perform and determining the value of the results relative to associated risks or financial costs’ (94). This implies that there is some logical or temporal sequence to a series of diagnostic investigations and tests that needs to take places in response to any presenting patient case.

I previously discussed the use of probabilistic reasoning that underpins development of CPRs. I have argued that more recently developed definitions of EBM are explicit about the use of such mathematical models. The application of computable forms of CPRs and probabilistic reasoning on their own is not sufficient for the purpose of providing diagnostic decision support. This point was well made in an on-line discussion of the merits of the Heneghan paper when the contributor said: ‘The appropriate model for diagnosis is not Bayes theorem alone. Another ‘theorem’ is also needed to explain how differential
diagnoses are formed from presenting complaints or other triggers (also called ‘diagnostic leads’) and how other information are used to differentiate between them (also called ‘differentiators’). The arithmetic of this process needs to be understood in order to practice evidence-based medicine properly’ (95).

In order to understand how CPRs may be potentially applied as a tool in clinical practice it is useful to place their use in a broader diagnostic strategy and context. As previously described, a clinician needs to formulate and consider the evidence for all possible differential diagnoses when a patient first presents with a particular clinical complaint. This is done by considering each differential diagnosis and can involve ‘ruling out’ differentials based on the underlying diagnostic cues as presented by the patient.

Whilst CPRs are rarely used to make a positive diagnosis they are a useful tool for excluding potential diagnoses that are considered as potential differentials. In the case of any specific patient case, suitable CPRs relating to the differential diagnoses under consideration may indicate the absence of certain conditions that are then considered as ‘rule outs’ (23). Their appropriate use can be applied as a tool to reduce the possibility of diagnostic error at the outset through identification of the correct possible differentials to consider (23, 70, 71). The appropriate use of ‘rule outs’ also acting as an important patient reassurance function of the consultation is usefully expressed in this comment relating to the Heneghan paper describing the diagnostic lifecycle: ‘General Practice consultations should, in my opinion, conclude with a statement about what is not the problem (often allaying worrying ideas/concerns/fears). Making a diagnosis is rare in unscheduled GP visits; this should be acknowledged’ (96).

2.7 Translating Evidence into Practice – the CPR as a Broader Expression of Epidemiological Knowledge

An awareness of the importance of the characteristics of patient populations and the clinical context within which they exist was established in its earliest forms as colourfully expressed by Hippocrates, one of the founding fathers of medicine:
'Whoever wishes to investigate medicine properly should consider the mode in which the inhabitants live, and what are their pursuits, whether they are fond of drinking and eating to excess, and given to indolence, or are fond of exercise and labour' (97).

Any technology based vision for CPR development must support a broader clinical context; the translation of clinical knowledge obtained through observational epidemiology in clinical practice. It is important therefore to ensure that any technological development of CPRs is still firmly grounded within the broader accepted definitions of epidemiology and that we capture any population and clinical contexts explicitly in our model representations. The application of technology to the problem does not in any way change these accepted definitions and concepts that are inherently part of epidemiological practice. Modern definitions of epidemiology are useful in making this fact explicit within the context of this research:

'Epidemiology is the study of the occurrence and distribution of health-related states or events in specified populations, including the study of the determinants influencing such states, and the application of this knowledge to control the health problems' (98).

I have discussed the limitations regarding the potential wider applicability and validity of CPRs beyond the original derivation population from which they were produced. This highlights an important consideration that is not always explicitly stated in all CPR research literature. CPR tools must be derived from and therefore exist within the confines of a specific population and clinical context. The application of any particular CPR may or may not be more widely applicable.

A CPR that has been narrowly validated and performs well within a particular population and clinical setting may be wholly inappropriate outside of that context. As such they must be considered to be specific fragments or instances of broader epidemiological knowledge where the clinical and population context is hugely important and is established through the conduct of a form of observational epidemiology on a well-defined population.
2.7.1 CPRs as a Method of Reflecting Clinical Observations

The traditional gold standard mechanism for conducting population based epidemiological research is the cohort or cross-sectional study. This typically involves recording observations over time without any intervention on a defined cohort population with a view to empirically establishing contributing or excluding those factors that are associated with a specific clinical outcome. This is elaborated on by Saracci:

‘Within epidemiology, observational studies are by far the most common. Epidemiologists observe what happens in a group of people, record health-related events, ask questions, take measurements of the body or on blood specimens, but do not intervene actively in the lives or the environments of the subjects under study’ (99).

Manually intensive recruitment of patients to cohort studies have traditionally been the mechanism used to derive CPRs but this may begin to change as research emerges that suggests alternative technology based approaches (100). From a technological perspective the ICT based solution that will be proposed here is still a form of observational epidemiology. I will investigate in the following sections how an infrastructure might be provided that allows us to replace traditional cohort study based observation with a process that exploits aggregated electronic sources of patient data as a basis for providing an electronic cohort of observations, as a viable and more efficient alternative to the traditional epidemiological observational study.

2.7.2 Translational Medicine

Whilst the concept of practicing evidence-based medicine tends to focus around providing clinicians with tools and formal processes that facilitate access to the latest available evidence on which to base their decisions, there is another critical element that is problematic. Where will the latest evidence-based data upon which clinical decisions will be made come from and how do we ensure that it is up to date and accurate? Currently evidence is generated from observational data obtained from cohort studies or patient registries and synthesised into research guidelines. Studies have demonstrated that as clinical guidelines are refined through research efforts, there is a considerable time
delay between when that new research is made available and when frontline clinicians begin to actually apply it in practice (101-103).

The currently accepted gold standard for assessing the impact of derived guidelines deployed in clinical practice is the randomised control trial (RCT). There may be a considerable period of time when practicing clinicians are actually out of step with the latest available research as generated through research studies relating to a particular topic. The study of the pathways and mechanisms that enable clinical research knowledge to be translated into actual clinical practice is known as 'translational research'. The term 'translational research' can be defined as follows:

- ‘research that seeks to characterize the sequence of events through which a scientific discovery moves between basic scientists, clinical researchers, practitioners, and consumers, and to find more effective ways to facilitate this process’ (104).

The translational process can be further broken down into distinct phases where two problematic areas or 'gaps in translation have been identified (105):

- type 1 translational gap – this relates to delays in moving from basic scientific development of theories to more structured clinical trials that establish more rigorous knowledge discovery
- type 2 translational gap – this relates to delays in applying knowledge into clinical practice to establish the impact of new knowledge and assess its actual effectiveness on clinical outcomes

I conclude that any proposed solution that wishes to provide for effective practice of evidence based medicine should be strongly linked with translational research by focussing on knowledge generation (type 1 translation) and knowledge dissemination in practice (type 2 translation). This requires investigation of more efficient mechanisms to quickly translate research data into actual clinical knowledge along with mechanisms to make that knowledge available to front line clinical practitioners to assess its impact in clinical practice. Any proposed solution must therefore address these issues through utilisation of electronic models of both research and clinical evidence data to
support more efficient ways to disseminate this knowledge between research and frontline clinical environments.

2.8 Clinical Literature Review Conclusions

The development and documentation of CPRs for use in primary care practice has traditionally been done in a traditional literature intensive way. This has held back their widespread use as a useful diagnostic tool that can be deployed and applied as part of wider recognised diagnostic strategies. Development of CPRs has largely focussed on the ‘derivation’ phase with the subsequent phases of validation and impact analysis being neglected.

Traditional CPR development has focussed on a narrow epidemiological interpretation of CPRs. The CPR lifecycle can be summarised as:

- derive small numbers of CPRs focussed on explicitly defined but narrow patient populations using traditional observational cohort studies
- test the validity of these CPRs on the original derivation population
- test the wider validity of these CPRs against other patient outside of the original derivation population
- where narrow or broad validation fails, go back and repeat the derivation process again on newly defined patient populations that may require new research studies.

I will refer to this as an ‘optimistic’ approach to CPR validation which makes optimistic assumptions as to the potential wider generalisability of limited findings obtained from the derivation phase of the CPR lifecycle.

The growing wealth of electronic patient data provides an opportunity to use ICT solutions to refine this lifecycle and make it more dynamic and efficient. In the ICT literature review described in the next section I will investigate ICT based solutions that will allow us to move towards what I will describe as a ‘targeted’ approach to CPR validation. Broadly speaking the desired targeted approach to CPR validation can be expressed as:
• utilise large volumes of aggregated electronic data to derive large numbers of CPR variants focussed on many explicitly defined patient population variants
• represent the CPR variants as models that explicitly capture the clinical and population context each one applies to
• test the applicability of these CPRs against other suitable patient populations using EHR patient data contexts as a trigger for decision support tools.

This approach represents an attempt to move from ‘optimistic’ CPRs to ‘targeted specific’ CPRs for any particular patient presenting case.

The technical literature review section that follows will review the strengths and limitations of previous research efforts that have either conceptually described or actually implemented such a practice.
3 State of the Art – Electronic Application of CPRs

3.1 Technical Literature Review

In order to propose a suitable electronic framework for deriving and implementing electronic CPRs we must understand both the successes and the limitations of previous work carried out in wider but related technical research areas. In its recommendations for improving the diagnostic process in healthcare, the Institute of Medicine highlighted the role that ICT can play. It importantly highlighted the broader multi-disciplinary considerations beyond the domain of ICT that require our attention:

‘Recommendation 3a: Health IT vendors and the Office of the National Coordinator for Health Information Technology (ONC) should work together with users to ensure that health IT used in the diagnostic process demonstrates usability, incorporates human factors knowledge, integrates measurement capability, fits well within clinical workflow, provides clinical decision support, and facilitates the timely flow of information among patients and health care professionals involved in the diagnostic process’ (54).

The purpose of the literature review therefore is to:

- identify barriers and limitations of previous development of diagnostic decision support tools
- identify suitable technologies for building and implementing our prototyping solution
- understand the emerging work that describes the goals and development of a learning health system
- identify the legislative implications that should be considered of implementing eCPRs for diagnostic decision support.
3.2 Diagnostic Decision Support Systems

3.2.1 The Role of Clinical Decision Support Systems

Decision support systems have a long and sometimes controversial research history (106, 107). Many different definitions identify the characteristics of a clinical decision support system. The following definition captures the core concepts and stresses their relationship to the practice of evidence-based medicine, providing a suitable context for this research: a clinical decision support system is ‘software that is designed to be a direct aid to clinical decision-making, in which the characteristics of an individual patient are matched to a computerized clinical knowledge base and patient-specific assessments or recommendations are then presented to the clinician or the patient for a decision’ (108).

The exact nature of the ‘patient-specific assessments or recommendations’ and the delivery mechanism used to present that information to the patient or clinician can vary greatly (94). This has resulted in a number of different ‘types’ of clinical decision support system that address particular clinical areas such as:

- computerised physician order entry
- appropriate medicines management
- diagnostic aids
- triggered alerts and reminders highlighting tasks to be done or potential conflicts in clinical information recorded
- full electronic implementation of clinical guidelines (94, 109).

Their demonstrable efficacy in clinical practice however has been limited. One reason is that research impacts of implementing CDSS have frequently been assessed as a technical driver of process change. Ideally they should more usefully demonstrate a measurable positive impact on practitioner performance that leads to directly attributable and measurable improvement in patient outcomes (110). But more promising results have been demonstrated in research environments outside the clinical area of diagnostics (111-113).
3.2.2 The Challenges of Implementing and Assessing Diagnostic Decision Support Systems

In the specific context of diagnostic decision support, the logical implication of the conclusions by Garg et al., is that assessment of positive impact on patient outcomes should be established through conduct of a clinical trial (110). The clinical trial aims to show that a technical intervention measurably improves diagnostic accuracy of clinical practitioner’s performance when compared against another gold standard reference diagnostic test or process. The diagnostic process as described in section 2.3.4 highlighted the fact that improvements in diagnostic accuracy can result in more timely and appropriate follow-up treatment and care. An increase in diagnostic accuracy therefore can reduce diagnostic error and improve patient safety by reducing unnecessary delays or treatments that have been shown to adversely impact on patient prognosis in serious illnesses like cancer (114). This can only be demonstrated by observing subsequent related patient outcomes, a task that has proved very challenging in practice (110).

Traditional approaches to diagnostic decision support have lacked broad acceptance for a number of other well documented reasons: poor integration with EHRs and clinician workflow, static black-box rule based evidence that lacks transparency and trust, usage of proprietary technical standards hindering wider interoperability (93, 111, 115-117). Despite these problems there is an increasing recognition of the need to realise the potential value of implementing decision support systems more generally. This is reflected in their inclusion as important components of wider government ICT based health policy legislation in practice (118-121).

3.2.3 The Evolution of Diagnostic Decision Support Systems

The evolution of clinical decision support development reflects attempts to tackle the problems previously described. Wright et al. have categorised the chronological development of clinical decision support systems, shown in figure 3-1, that address workflow and integration issues, interoperability standards and also separation of the knowledgebase as a separate service distinct from the tools themselves (122).
Wright identified four key phases in the development of decision support systems:

- **standalone systems** offered decision support in proprietary standalone applications that were capable of any integration with other clinical applications.
- **integrated systems** were capable of some integration with existing clinical systems in a limited clinical environment but were not standardised and therefore not shareable beyond that clinical context.
- **standards-based systems** applied recognised standards to allow sharing and interpretation of clinical data outside of the original applications in which they were derived or used.
- **service model based systems** disconnected the underlying knowledgebase from the decision support tool and made it available to potentially many client applications via standardised interfaces accessible over the internet.

The evolution of diagnostic decision support systems has largely focussed on implementations of what can be described as ‘diagnostic symptom checkers’.

Figure 3-1 Key phases in the evolution of clinical decision support systems - Courtesy of Wright et al. (122)
These implement rule based knowledgebases for diagnosis with limited degrees of clinical acceptance because of many of the same limitations identified in general decision support systems previously described (107).

Based on the 4 phased trends that can be identified in the evolution of general decision support tools up to 2007, this research on diagnostic decision support should at a minimum support those desirable features in its implementation: workflow integrated, standards based, service-oriented. This work will therefore identify further desirable features that have emerged since 2007 to extend this to a ‘phase 5’ architecture to implement diagnostic decision support. In order to do this we must examine other attempts to provide diagnostic decision support using eCPRs.

3.2.4 Implementation of Diagnostic Decision Support Systems Utilising eCPRs

For clarity it should be stated that this pertains specifically to diagnostic decision support systems that implement the formal CPR structures and threshold approach to risk estimation as previously described in section 2.6. A number of studies that describe attempts to implement eCPR based diagnostic decision support have been highlighted by El Kareh (115).

These numbers are limited when identifying implementations of eCPRs integrated into the actual clinician workflow in primary care using the EHR as a trigger for diagnosis (12, 26, 123). Kharbanda describes implementation of an EHR integrated eCPR for diagnosis of appendicitis in children using an implementation based on the paediatric appendicitis score (124). The QRISK and Framingham scores for estimating cardiovascular risk have been widely implemented with EHR systems (125-127). The fact that few eCPRs have been integrated into clinician workflow is a recognised barrier to implementing effective decision support more generally and as such is reflected in the lack of evaluation of clinical impact of these systems in practice (93, 117, 128).

Few studies describe eCPRs that have been clinically trialled to assess their impact on clinical outcomes in ambulatory care clinical practice (12). This is consistent with more general findings that highlight traditional CPRs have not
been developed in primary care to the point of assessing impact analysis (34). McGinn describes an implementation of the Walsh rule for diagnosing streptococcal pharyngitis and the Heckerling rule for diagnosing pneumonia as part of an integrated EHR. A subsequent clinical trial using the eCPRs as an intervention group demonstrated the clinical impact of the eCPRs in ruling out potential diagnoses and thereby reducing unnecessary antibiotic prescribing and lab testing with associated cost savings (129).

As seen with the examples of EHR integrated eCPRs, the majority of eCPR implementation studies focus exclusively on implementations of individual CPRs that target diagnosis of a particular clinical condition, as opposed to a more general implementation of eCPRs that could be used to represent and deploy them more generally to diagnose any clinical condition (93, 117, 128). Douma describes an effective implementation and validation comparison of 4 eCPRs in secondary care for diagnosing pulmonary embolism (130). Pierre-Marie describes an implementation of the Geneva score for diagnosing pulmonary embolism using hand-held devices (131). The use of eCPRs for identifying bacteraemia in blood samples was demonstrated by Wang (132). As such it can be stated that there is a significant amount of technical and modelling rework is taking place to implement each individual eCPR in each single organisational system (133).

A study implementing the Ottawa Ankle rule as an ICT based decision support intervention highlighted a number of organisational barriers that need to be addressed to successfully implement eCPRs at the point of care including:

- need to educate and promote organisational awareness of the intervention and delivery mechanism
- requires clinical agreement and consensus at an organisational level for the intervention and delivery mechanism
- a case must be made for financial and clinical benefits of the intervention on patient outcomes
- organisational resistance to new ways of clinical working must be overcome (134).
These attempts also demonstrate some barriers unique to CPRs that need to be considered when implementing eCPRs as part of decision support systems:

- validation/impact analysis of eCPRs is restricted due to lack of connectivity to wider patient populations beyond the original electronic tools that they are initially deployed and tied to (12, 26, 33)
- with some exceptions where evidence is disseminated using open standards and separate from the application itself (27, 135, 136), decision support tools are tied to specific proprietary clinical systems, which lack support for wider Electronic Health Record (EHR)/workflow integration across other systems (124, 130-132)
- implementations of decision support tools focus on implementing models that are unique to individual instances of eCPRs and are not easily generalisable to implement other eCPRs, resulting in redevelopment effort for each implemented rule (12, 27, 124, 130-132, 137)
- successful implementations and organisational acceptance of the deployed tools necessitates a collaborative multi-disciplinary approach to define the nature of the intervention required and actual workflow of the eCPR in practice (12, 54, 134).

These limitations can be considered part of the wider problem of using ICT tools to provide more effective translation of clinical research knowledge into clinical practice. This will be discussed in more detail in section 3.6 as a key driver for development of what has been described as the ‘learning health system’.

3.2.5 Workflow Implications of Diagnostic CPRs

As previously discussed, although all CPRs share a common structure they may have different clinical applications related to diagnostic or prognostic outcomes. The type of CPR has significant implications for clinical workflow integration. A diagnostic CPR estimates the probability or risk of the presence or absence of a disease at a fixed point in time, for a specific individual (91). A prognostic CPR is more complex having an additional temporal aspect after the prognostic prediction has been made.
Validation of a prognostic CPR requires follow up to see if a particular clinical event relating to the prognosis subsequently transpires at some defined time window into the future (138). The workflow implications are less complex to embed diagnostic eCPRs within EHRs or decision support applications, since the diagnostic eCPR can be event driven and trigger a recommendation made and recorded at a fixed point in time without the need for future follow up.

On the basis that we are focussing on diagnostic eCPRs only, interoperability considerations have been a core focus for development of this model before considering more complex time dependent workflow integration. Prognostic eCPRs will not be considered in the scope of this work.

3.3 Rule Based Versus Model Based Diagnostic Decision Support

As we have seen, Wright et al. described a four phase evolution of clinical decision support tools up to 2007. These approaches to implementing decision support have relied on a knowledgebase defined as a series of rules in the form of a database of knowledge facts. These may be triggered or combined together in the form of guidelines based on statements using a knowledge rule languages or rule engines such as Arden Syntax, GLIF and GELLO (133, 139-143).

A more recent approach to implementing a diagnostic knowledgebase is to represent knowledge in the form of a underlying model that explicitly defines the relationships between facts in a unified structure that more accurately represent the real world. Proponents argue that a number of desirable benefits are inherent in this approach (144):

- the assumptions and omissions inherent in the knowledgebase are more transparent as the model needs to reflect the particular process that the knowledge will be applied to (in this case diagnosis).
- It opens evidence up for wider scrutiny compared to a ‘black box’ approach
- models are independent of data and can still be developed without large volumes of required data
model based approaches provide for integration of multiple datasets into a single ‘epistemic framework’ that allows inference mechanisms to be applied to generate new knowledge

the very process of carrying out modelling may promote multidisciplinary discussion of the problem with domain experts which is beneficial in itself

easier programming maintenance through a cleaner separation of the knowledge representation itself from the syntax or programming logic that is used to apply it in practice.

The decision support development chronology proposed by Wright et al. can therefore be extended to bring it up to date. A fifth evolution can be added to the chronology, in the form of a trend towards model-based approaches to knowledge representation for the purposes of clinical decision support. The use of ontologies and what is more widely termed ‘semantic web’ technologies has been a growing area of model-based research and will be discussed in more detail.

3.4 Model Based Approaches - Semantic Web Technologies and Ontologies

The origin of the term ‘ontology’ is a philosophical one closely related to metaphysics. Philosophers such as Plato attempted to structure both the real and ‘enlightened’ worlds in hierarchical structures of related conceptual knowledge where ultimate enlightenment was attained at the upper levels of the hierarchy (145).

The term has more recently been appropriated for the purposes of describing a model based technique for representing knowledge domains for use by ICT systems (146). The development of ontologies has spawned wider research into what are known as the application of semantic web technologies. These technologies have received widespread research attention in recent years. The simple data representations used in the semantic web have been seen as a means of solving interoperability problems in the clinical domain, due to its successful application in other commercial knowledge domains (146-148).
An ontology provides a formal definition of a model that describes the core concepts and directionally named semantic relations that exist between those concepts. The ontology describes a selected knowledge domain of which defined questions can then be asked. The ontology describes a formal contract for meaning that can be used to formulate queries of data from the selected knowledge.

The application of what are termed ‘semantic web’ technologies in the form of ontologies has shown itself to be highly successful in achieving this in real world applications in areas as diverse as biomedicine, social networking and online retailing by providing for a common definition of the core concepts contained in a knowledge domain that can easily be shared and used by other applications outside of the original development application.

The use of an ontology provides for a clear abstraction of the definition of knowledge from the underlying operational applications that will make use of them. As an example, a clinical evidence ontology can be developed to be distinct from any diagnostic algorithm that will make use of the concepts described in that ontology. Because of this abstraction and the fact that the technologies used to describe ontologies are based on openly available recognised standards (defined by the world wide web consortium W3C), ontologies can then be re-used by other applications outside of their original application by providing mappings to their own repositories of application data using a single consistent interpretation of that clinical knowledge (149-152).

A primary reason that ontologies have been successful in other commercial applications is the simplicity of the underlying data representation used. This is in stark contrast to other models developed which have embedded a higher degree of complexity into the models and the associated data representation. The primary data representation mechanism underpinning ontology development is a standard known as 'resource description framework' (RDF) (150). The focus in RDF is to reduce knowledge description to an extremely
simple representation using statements in the form of ‘triples’. These define statements in the form of three component parts: subject-predicate-object. Looking at an example of triples in the clinical evidence domain, we may define generic concepts such as ‘Diagnosis’ and ‘DiagnosticCue’.

Another key difference is that relationships are directional in nature with pairs of inverse relationships allowing queries to be executed in either direction. We may define two inversely related directional relationships between these two concepts called ‘hasCue’ and ‘isCueOf’. Within each generic concept we can define concept instances that represent specific cases of those generic concepts that we wish to define for our application. As shown in table 3-1, we can then establish directional relationships between those instances that define triple statements.

Table 3-1 Example of triples and inverse relationships in the diagnostic domain

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Dysuria’</td>
<td>‘isCueOf’</td>
<td>‘UrinaryTractInfection’</td>
</tr>
<tr>
<td>- (DiagnosticCue)</td>
<td>– (Relationship)</td>
<td>– (Diagnosis)</td>
</tr>
<tr>
<td>‘UrinaryTractInfection’</td>
<td>‘hasCue’</td>
<td>‘Dysuria’</td>
</tr>
<tr>
<td>– (Diagnosis)</td>
<td>– (Relationship)</td>
<td>– (DiagnosticCue)</td>
</tr>
</tbody>
</table>

By building a complete set of these triple statements we define both the generic concepts and relationships that exist and also the specific instances of those concepts that we wish to work with for our particular application. Collectively these define what we know about the knowledge domain we are modelling.

The focus on ontology definition is concerned with describing the semantic characteristics of the relationships that exist between the defined ontology concepts in a more complete way than traditional data modelling techniques such as Unified Modelling Language (UML) does (154). In UML relationships between concepts are bi-directional and generally restricted to defining cardinality between concepts. In an ontology, relationships are defined in a
single direction along with a definition of characteristics that describe the specific semantic nature of those relationships and how they can be applied to the generic concepts defined in the ontology. This is very suited to defining characteristics of clinical domains by allowing data to be queried in a directional manner from top-down or bottom-up.

As a diagnostic example I can start from a diagnosis working down to diagnostic cues, or starting with a diagnostic cue and working up to a diagnosis. The ability to achieve this semantic expression is achieved through use of additional metadata technologies, such as ‘RDFS’ (RDF-schema language) and ‘web ontology language’ (OWL) (151, 152). These are layered upon the basic data representations of RDF specifically with a view to defining generic formal models that fully describe a particular knowledge domain in the form of an ontology.

The lack of generalisability of previous attempts to implement electronic forms of CPRs suggests that a formal generalised model of eCPRs can contribute one part of an overall solution. This generalisable model needs to capture the conceptual components needed to model any eCPR rather than just specific instances of them as described in section 2.6. The review of the eCPR literature described in section 3.2.4 identified a limited number of implementations of eCPRs and highlighted the fact that these implementations focussed on specific individual eCPRs rather than a more general approach to representing and deploying eCPRs. Where a more generalised approach to eCPR development was done, it was done using a rule based approach, as described in section 3.3, rather than a model based approach to representation (27, 135, 155, 156). On that basis I conclude that such a general model based eCPR representation does not currently exist.

The representation of CPRs electronically in isolation will not solve the problems I have previously described in relation to traditional CPR development or their application in decision support systems. A translational approach to evidence generation is needed where a suitable electronic infrastructure can both host the models for the purposes of both evidence generation and evidence dissemination. The application of ICT can then be used to move
traditional literature based evidence further up the pyramid of evidence by facilitating faster evidence generation, dissemination and validation in practice.

3.5 Where Traditional CPRs Fit into ‘the Pyramid of Evidence’
The traditional literature based lifecycle of CPR development previously described is considered to be at a low level of technological development with respect to what has been termed ‘the pyramid of evidence’ (92, 157). In order to address the limitations previously described we need a technology focused vision of an eCPR development process that can utilise the potential aggregated sources of electronic health records (EHRs) to move eCPR development to the top of the pyramid of evidence. One such vision that may provide a suitable technological context and environment within which to discuss and achieve this is what has been termed ‘the learning health system’. The infrastructure that describes the LHS is still emerging and is highly ambitious in scale defining a number of core requirements to be in place.

3.6 The System Level Objectives of the Learning Health System
There are difficult challenges accessing and consolidating distributed data sources for research. The level of consolidation required may be at an organisational or regional level. More ambitious visions may see consolidation at a national or international level to facilitate development of national clinical guidelines for example. These distributed sources of potential research knowledge may all use different technical standards that don’t directly communicate with each other. Some approaches have employed a traditional data warehouse to move data into a centralised reporting location using an ‘extract, transform, load’ process (ETL) (32). Communication is typically implemented as a batch process rather than providing access in real-time. In addition movement and storage off-site of sensitive clinical data requires dealing with potentially complex cross-institutional policies on data privacy and protection (158). Is there a way to access this data from its secure location in a pluggable, connected and unified way? Ideally data would be directly accessible and linkable in place from the source. The challenge is to provide real-time connectivity and linkage of heterogeneous data sources for the purposes of research and evidence generation when we need it.
This is a primary objective of what has been called ‘the learning health system’ (29, 159). The concept of the LHS is an increasingly important research topic in addressing these challenges. The LHS is strongly linked to the goals of translational research, the study of the pathways and mechanisms that enable clinical research knowledge to be translated into actual clinical practice. The nature of LHS research is necessarily multi-disciplinary as it requires an ambitious vision requiring technical and policy development.

3.6.1 The LHS ‘Virtuous Cycle of Health Improvement’

A core function of the LHS is to provide a complete distributed infrastructure that is ‘capable of engendering a virtuous cycle of health improvement’ (28). The primary functions of the LHS provide a translational system that serves two complementary objectives: the generation of clinically valid knowledge, and the subsequent electronic curation and translation of that knowledge efficiently into clinical practice for the purpose of providing clinical decision support. This requires the integration and management of three distinct types of health knowledge:

- routinely collected health knowledge (e.g. EHRs)
- research study knowledge in the form of electronically conducted research trials or cohorts of observational data obtained from electronic disease registries
- derived clinical knowledge through the application of data mining and evidence discovery techniques applied to both routine and research knowledge (30).

The LHS concept consists of a distributed electronic infrastructure supported by a strong emphasis on required policy instruments to enable federated use of large scale sources of clinical data dynamic evidence generation and dissemination (28, 160). This requires consideration of diverse policy-driven topics such as data privacy and protection, medico-legal responsibility for generated evidence and the concept of treating ‘software as a medical device’ when implementing decision support.
The concept has moved beyond the realms of theory through research initiatives in both the US and Europe that are developing and testing functional implementations of the core components that constitute the LHS (159, 161-163). These research initiatives have defined what should constitute the core components of the LHS (159, 161-163). They describe high-level requirements and capabilities needed to satisfy the research vision of the LHS.

Within the 'virtuous cycle of health improvement' a number of characteristics have been identified as requirements to support this knowledge translation capability including:

- generating valid clinical knowledge
- packaging and curation of knowledge so it is widely accessible and actionable and putting knowledge to use to effect change
- development of meaningful use of the EHR to support diagnostic and therapeutic support based on evidence
- development of a computable representation of research evidence made available to EHR systems as a web service
- develop a means of providing diagnostic or therapeutic prompts within an EHR that works across a variety of EHR systems
- enabling cyclic updates of clinical knowledge through provision of an evidence feedback mechanism through which updated observational clinical data can be fed back into the knowledge generation process (28).

3.6.2 The eCPR Development Lifecycle as Part of the LHS

The LHS represents an attempt to support both evidence discovery and subsequent evidence dissemination, using a single infrastructure connected to a distributed network of data sources. The application of an LHS infrastructure to the specific needs of eCPRs has the potential to support derivation and dissemination of this form of eCPR based clinical knowledge in practice.

The traditional form of dissemination of CPRs has been largely literature based which has been a limitation for their broader uptake. I have considered modelling approaches that can facilitate the electronic representation of eCPRs.
This is just one part of the ‘virtuous cycle of health improvement’ that was previously discussed. In order to deliver such an infrastructure in full we must also consider electronic derivation of eCPRs as part of that process. Data mining routinely collected health data in the form of the EHR provides one mechanism through which this may be achieved. Data mining on its own however does not constitute the LHS. It is the combination of data mining to discover knowledge and its subsequent dissemination using a suitable model based representation that defines the core characteristics of an LHS.

3.7 Data mining eCPRs from the EHR

3.7.1 Observation, Measurement and Evidence Discovery - the Generic Data Mining Process

Data mining emerged as a significant topic of interest to both research and industry in the mid 1990’s (164). It can be argued that the basic concepts behind data mining are not something new in itself but that ICT enabled it on a greater scale. Advocates for evidence discovery based on observation and measurement of living things can be traced as far back as Aristotle and David Hume (165, 166).

The application of ICT to this task has implemented more efficient means and methods used to analyse the data, enabling data analysis on a quicker and bigger scale than could previously be achieved. The growing body of research applying data-mining across many industries resulted in attempts to describe a more generic, standardised and reusable data-mining process. One of the most successful and widely applied data-mining methodologies is the ‘CRoss Industry Standard Process for Data Mining’ (CRISP-DM) (167, 168). The basic steps in data-mining for knowledge discovery include variations on the following identified CRISP-DM steps (167, 169):

- business and data understanding (data selection)
- data preparation (data pre-processing, cleaning and transformation)
- data modelling construction (data-mining)
- data model evaluation (interpretation/evaluation of results)
- data model deployment (application of knowledge).
This methodology has also been successfully applied to many different healthcare settings (164, 168, 170).

3.7.2 Structured and Unstructured Data Mining

The application of data mining in healthcare has broadly speaking spawned two distinct technical approaches to the task based on the type of data being analysed. For the purposes of this research we will consider data to be structured or unstructured:

- structured data is semantically well defined through the use of a chosen clinical coding scheme and usually stored in a well defined data representation format e.g. coded patient data held in an EHR (171-173)
- unstructured data is not semantically well defined and may simply be expressed as textual data from which semantic context or meaning needs to be further extracted e.g. free text doctor notes recorded in a patient record, or health related twitter messages as tweets (164, 174).

The steps to carry out data-mining on unstructured data can still fit the generic methodology as previously described but typically the data preparation step will be more complex involving additional manual textual processing including (174, 175):

- construction and definition of text interpretation rules from a training data set that allow appropriate keyword classification of textual data using either an existing or custom defined terminology related to the knowledge domain of interest
- target dataset analysis and deconstruction of unstructured textual data to identify key textual phrases of interest
- application of developed rules to do textual classification and semantic interpretation of textual phrases found in the target.

Free-text data can provide supplemental knowledge typically derived from descriptive results interpretation which is of value when used in combination with structured approaches (176). For the purposes of this work the focus will be exclusively on the case of knowledge discovery through structured data mining techniques applied to the EHR. This has been done for both practical
and scoping reasons. As will be described in detail later, this research has applied data mining techniques to data obtained from an openly available database of anonymous patient data called TRANSHIS (177). This data is openly available for research purposes but provides only structured and coded patient data. Free-text data is not supported in TRANSHIS and has therefore not been considered.

As highlighted by Rasmussen, the use of structured data also provides a simpler case example to demonstrate eCPR learning (176). The defined semantic meaning of terms allows easier aggregation of data since there is more confidence that different patients can be grouped where the same individual codes intend the same clinical meaning. The unstructured data mining process as I have described it here requires additional computational steps. On that basis it has not been considered to provide a more manageable scope for this research work.

3.7.3 Data Mining for eCPRs and Clinical Decision Support

Data mining electronic sources of patient data is one mechanism that might be used as an alternative to the traditional CPR development lifecycle to discover ‘clinically valid’ CPRs. The process of data mining for primary care diagnostic eCPRs requires:

- identification and use of suitable well coded aggregated patient data that allows for secondary re-use
- detection of diagnostic relationships trends of interest in the coded data
- quantification of those diagnostic relationships in the form of eCPR tools.

The success of any such data mining approach is highly dependent on the identification of suitable high quality sources of electronic patient data to which we can apply it. Rich aggregated electronic sources of family practice data currently exist in the United Kingdom (Clinical Practice Research Datalink (CPRD), The Health Improvement Network (THIN)), Canada (Canadian Primary Care Sentinel Surveillance Network), Netherlands (NIVEL), and Ireland (Irish Primary Care Research Network (IPCRN)) (178-182). The IPCRN is a
collaborative research network that allows Irish GP practices to submit anonymised coded patient data relating to medication prescribing obtained from local practice EHR to facilitate centralised medication research studies. In return the local practice can make use of a number of practice level reporting tools relating to medication management, cardiovascular management, so that GPs can gain analyse their own patient practice population to facilitate practice medication prescribing safety and reporting (182).

The use of aggregated sources of healthcare data for secondary use has sometimes been controversial though with concerns expressed relating to patient privacy and access by commercial interests (158). Difficulties have also been expressed relating to ease of access to clinical data using third party EHR tools provided by potentially many different commercial providers. The issue of ‘information blocking’ in the health sector has recently been highlighted in a US government report identifying difficulties with openly accessing patient electronic health record data for the benefit of wider research (183). Information blocking relating to health care providers and health IT developers, including vendors of EHR technology is defined as ‘knowingly and unreasonably interfering with the exchange or use of electronic health information’ (183). The report concluded that information blocking was an ‘impediment to an interoperable learning health system’ and the findings suggested that it is a serious problem and one that is not being effectively addressed.’

An alternative data source for data mining has been to make use of data obtained from openly available EHR products that are not tied to any particular commercial interests. The TRANSHIS EHR is an example of an openly available EHR that can be freely be downloaded and used in multiple population sites for the purposes of gathering large volumes of aggregated electronic patient data for research (177). Research initiatives have demonstrated some success in analysing this data to detect diagnostic trends or patterns (173, 184, 185).

3.8 Decision Support Tools as a ‘Medical Device’

The generic data mining process that was previously discussed highlighted the need for a mechanism for deployment of the knowledge outputs derived from
that process. From a clinical context we saw that one of the methods for deployment of such knowledge is in the form of diagnostic clinical decision support tools. From a patient safety perspective the clinical reliability and quality that can be attributed to any electronically derived clinical knowledge must be considered when deploying it using decision support tools. As such quality assurance of the decision support tools and underlying knowledge used is a crucial issue to consider. This is even more critical in the case where that knowledge is to be used to treat patients in a diagnostic capacity.

In Europe and the US this fact has been recognised in legislation that treats software that has been developed for diagnostic purposes as a ‘medical device’ in its own right. This places consideration of quality assurance issues to the forefront for such categories of software.

The core pieces of European legislation to be considered relating to medical devices are:

- directive 90/385/EEC relating to active implantable medical devices (186)
- directive 93/42/EEC concerning medical devices (187)
- directive 98/8/EC concerning the placing of biocidal products on the market (188)
- directive 2007/47/EC concerning inclusion of software as a medical device (189).

### 3.8.1 EU Definitions

For the purposes of clarity it is useful to highlight the relevant definitions as defined in the legislative documents. The following definition of what constitutes a ‘medical device’ is taken directly from the text of directive 2007/47/EC:

“A ‘medical device’ means any instrument, apparatus, appliance, software, material or other article, whether used alone or in combination, together with any accessories, including the software intended by its manufacturer to be used specifically for diagnostic and/or therapeutic purposes and necessary for its proper application, intended by the manufacturer to be used for human beings for the purpose of: — diagnosis, prevention, monitoring, treatment or alleviation of disease, — diagnosis, monitoring, treatment, alleviation of or compensation
for an injury or handicap, — investigation, replacement or modification of the anatomy or of a physiological process, — control of conception, and which does not achieve its principal intended action in or on the human body by pharmacological, immunological or metabolic means, but which may be assisted in its function by such means” (189).

The inclusion of stand-alone software without the presence of hardware devices in this definition is detailed in the following supplemental statement contained in the directive text:

'It is necessary to clarify that software in its own right, when specifically intended by the manufacturer to be used for one or more of the medical purposes set out in the definition of a medical device, is a medical device. Software for general purposes when used in a healthcare setting is not a medical device’ (189).

3.8.2 EU Medical Device Classes

Directive 93/42/EEC legislation also describes the characteristics and impact of medical devices that could potentially constitute a risk to patient safety. The classification depends on rules that consider the medical device potential risk factors such as duration of body contact, its invasive character, its use of an energy source, its effect on the central circulation or nervous system, its diagnostic impact or its incorporation as a medicinal product (190).

The classification of medical devices is therefore categorised in four broad classes that quantify the potential risk as low (class I) to high (class III). Where the medical device directive legislation applies, all classes other than certain sub-classes within class I are subject to third party verification and certification of legislative compliance by an accredited ‘notified body’ (NB) (as described in the subsequent section) (191). Certain sub-classes within class I devices (least risk) can be self-certified on demonstration of compliance by just the manufacturer themselves.

The four medical device classes are (190):

- class I – low risk
• class IIa – medium risk
• class IIb – high risk
• class III – highest risk.

In the context of diagnostic eCPRs it was considered that this falls into the category of class IIa on the basis that there is a diagnostic element to the software, but the risk is medium rather than high, because the software is not active (i.e., worn all the time) or invasive (inserted into the body). The implications of this are that if the medical device directive were to be implemented then it would be subject to third party verification of compliance. If the tools being developed were to be deployed in multiple EU member states then that certification is subject to verification of compliance by the notified body selected from each member state in which deployment will take place.

3.8.3 Compliance Procedure

It is the responsibility of each EU member state to enact EU legislation relating to medical devices by appointment of a ‘competent authority’ (CA) responsible for certification of medical devices (192). In Ireland for example the responsible CA is the Health Products Regulatory Authority (HPRA) (193) and in the UK the responsible CA is the Medicines and Healthcare products Regulatory Agency (MHRA) (194). The CA acts on behalf of the government of each member state to ensure that the requirements of the medical device directives are reflected and enforced in national law. The CA only has jurisdiction within the member state within which it was appointed.

All certified medical devices must be identified in the EU with the Conformité Européenne mark (CE mark) on their packaging (195). Whilst the authorisation of medical devices is guaranteed by a ‘Declaration of Conformity’ issued by the medical device manufacturer itself, it is also subject to third party verification by a Notified Body (NB) (for all device classes except a subset of class I). A Notified Body is a public or private organisation appointed within a member state that has been accredited to validate the compliance of the device to the applicable standards associated with the EU legislation as enacted by the CA for that particular member state. Each member state has its own list of
recognised NB’s and in Ireland the NB is the National Standards Authority of Ireland (NSAI) (196).

3.8.4 **EU Applicable Standards**

In order to demonstrate legislative compliance and certification, best practice is to implement the recognised associated International Standards Organisation (ISO) standards that are applicable to this area. These standards define requirements for general design and manufacture quality procedures, management of risk in relation to patients, and software development processes.

These are:

- ISO 13485 – quality systems relating specifically to the design and manufacture of medical devices (considered a subset of ISO 9001) (197)
- ISO 14971 – systems relating to risk management of medical devices (particularly with respect to potential risk of adverse events to patients) (198)
- IEC 62304 – systems relating to the software development lifecycle of software classed as a medical device (199).

It should be noted where additional third party products are also being used as building blocks in the context of overall software design and implementation that they are also subject to satisfying adherence to quality requirements e.g. use of third party open-source software components.

3.8.5 **Medical Devices Outside the EU**

In the US context the equivalent body with responsibility for medical devices is the US Food and Drug Administration (FDA). The relevant legislation is detailed in ‘Code of Federal Regulations Title 21- Part 820’ (200). They also define classifications of medical devices (Part 860) for operation in the US (201).

The FDA defines a medical device as ‘an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including a component part, or accessory which is (202):
recognized in the official National Formulary, or the United States Pharmacopoeia, or any supplement to them,

- intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or

- intended to affect the structure or any function of the body of man or other animals, and which does not achieve its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of any of its primary intended purposes.’

It is clear based on the FDA definition that software with the purpose of supporting diagnosis therefore falls into the category of a medical device.

3.8.6 Research Interpretation of Medical Device Directive

The key features as described in the medical device definition from the perspective of this research are that stand-alone software developed for the purposes of diagnostic support falls within the context of the medical device directive and as such there must be certain guarantees of quality for it to be used in real clinical practice. From this perspective it can be argued that the work in this research falls definitively within this context. I have previously discussed the need to improve patient safety by using ICT tools to improve diagnostic accuracy by addressing heuristic bias found in the diagnosis process in primary care. This is done with a view to reducing subsequent delays in treatment and care that have been shown to adversely impact on patient prognosis particularly in serious illnesses (114). It therefore follows that if these ICT diagnostic tools are implemented incorrectly or underpinned by poor quality evidence then this has the potential to negatively impact diagnostic accuracy, resulting in potential delays in diagnosis, appropriate care and treatment, along with a poorer patient prognosis.

The key aspect that differentiates this work from these definitions is that the ‘medical device’ is not actually being commercially developed with a view to immediate market application against real-world patient subjects. Therefore it can be reasonably stated that there is no potential risk to actual patient safety
within the context of the research work being done as real patients are not being used. I fully acknowledge however the real-world risks that could result for patients if these tools were to be employed in real clinical practice along with the importance of demonstrating an understanding of the implications of the medical device directives for this research.

For the purposes of scoping this research project however the directive will not be applied here as it would impose significant overheads on effort to achieve as demonstrated in the discussion of the certification process itself. I acknowledge that this is required at a future date to move this work from a research context to a real-world context where validation of deliverables is carried out on real patients. A useful secondary output of this work however is to consider what limitations are imposed by this directive on software based health research. It appears to have implications for the LHS that have not yet been considered in great policy detail. This will be addressed in the discussion later.

3.9 Technical Literature Review Conclusions

As we have previously seen, a practical problem exists when trying to effectively apply evidence based medicine in a time-pressurised consultation without supporting tools to do it. The application of ICT based interventions in the form of clinical decision support systems is one mechanism to complement heuristic 'top down' approaches to the diagnostic process with 'bottom up' analytical approaches that can be more quickly executed by computers (94).

The successful blending together of heuristic and analytical reasoning is dependent on the integration of these tools into clinician workflow (93). This effective integration of CDSS with clinician workflow requires development of new diagnostic strategies that address limitations in traditional clinical knowledge bases and reasoning. These should be deployed as part of one overall integrated strategy to succeed. The need to make use of data obtained from the electronic health record as a workflow trigger has therefore been highlighted as a key component for research in this area (50, 173). This has the potential to allow evaluation of the most up to date evidence available relating to any particular presenting patient case.
The provision of freely available electronic representations based on open standards of CPRs, according to our stated definitions, does not appear in the literature. There is scope therefore for development of tools that allow for easier dissemination, searching and use of CPRs along with facilities to record their actual application to clinical cases in practice with a view to accessing their validity and impact. A pre-requisite for such tools is the development of a suitable eCPR model to provide for an open standard based representation that facilities integration and interoperability with other clinical systems.

The more recent definitions of evidence based medicine that I have discussed suggest that developing effective CDSS tools also requires a model-driven approach. This requires the development of underlying data models for representation of empirically quantified clinical evidence. This has resulted in the consideration of technology-based interventions that may complement evidence-based medicine to improve patient safety through limiting diagnostic errors (44, 93, 203). This is an approach that is consistent with literature indicating that 'interventions directed at specifically encouraging both analytical and non-analytical reasoning have been shown to result in small, but consistent, improvements in accuracy'(64). In a study by Eva et al. (63), 60 clinical students diagnostically assessed electrocardiograms (ECG) and were split into two groups, one using exclusively non-analytical reasoning only, and the other using non-analytical combined with analytical reasoning. Out of 16 ECGs the mean diagnostic accuracy for the combined reasoning group was 8.13 compared to non-analytical reasoning group of 5.13 (p <0.01).

The ultimate vision is to provide for computable representations of CPRs that allow derivation, validation, dissemination, versioning and on-going revision from empirical sources of electronic primary care patient data. This can be complemented using extraction of patient cues and demographics from EHRs as a trigger for initiating appropriate rule execution. It can be stated at this point that the goal of this work is to define and demonstrate using a working implementation, an electronic framework that supports this vision for electronic derivation and dissemination of eCPRs in clinical practice.
The central role played by EHRs in primary care will only increase as legislative requirements, such as the ‘meaningful use’ initiatives in the United States, promote their wider use in the form of government backed ICT based health initiatives (118). These government based initiatives see EHRs as the workhorse for incremental development of increasingly sophisticated functionality that goes beyond basic patient chart recording and in effect turns them into ‘medical devices’ providing support in the form of clinical decision support. Decisions related to clinical care must still ultimately be based on the good judgement of the clinician with due consideration to the supporting evidence available.

The LHS provides a suitable infrastructure for development of a new breed of model based learning decision support infrastructure and tools. These tools can exploit the potential for appropriate use of the growing volumes of aggregated source of EHR data to support learning clinical evidence generation. The goal of implementing such an LHS is to describe and implement an architecture that is model-driven, service oriented, constructed using open standards, and supports a learning evidence base that utilises electronic sources of patient data. The architecture and implementation must support two critical aspects for a successful LHS: the model representation and translation of clinical evidence into effective practice, and the generation of curated clinical evidence that can be used to dynamically populate those models thus closing the LHS loop.

The scope of this work within the context of the CPR development lifecycle can now be stated. This work sits primarily in the stage 1 derivation stage of the CPR development lifecycle to enable derivation based on larger population samples. It can also address stages 2 and 3 through facilitating broader dissemination and enabling wider validation via access and triggering through the electronic health record. Stage 4 impact analysis of the development lifecycle will not be explicitly addressed and is outside the scope of this work as the assessment of clinical impact of a CPR represents a significant research effort in its own right.
4 Conceptual Framework for Development of eCPRs

The literature reviews have highlighted an opportunity to develop eCPR based decision support tools that can be integrated with the EHR and complement the existing diagnostic strategies employed by clinicians as part of their everyday practice. The goal is not to try to replace clinician judgement but to support diagnostic decision making by providing the right evidence available at the right time during the patient consultation to inform good decision-making. To do this we need evidence based tools that implement well defined guideline structures. As stated previously, these should explicitly allow for empirically quantifying the evidence being presented.

The literature reviews have also identified a number of key design features identified as best practice with respect to development efforts in decision support to date. I will develop this further by the addition of support for a model based approach with the inherent design advantages that it supports (93, 122). Our eCPR model should:

- support open source technical standards at all stages of development
- separate the knowledge-base from the actual application by making it a clinical evidence service accessible over the web
- represent clinical evidence in the form of a unifying ontology model that is independent of, but can be bound to, any desired clinical terminology or coding scheme
- support for direct two-way communication with third party EHR and clinical workflows.

4.1 Developing Smarter eCPRs as Part of the LHS

We have also seen that the general requirements of the LHS can be more specifically considered in the context of the specific requirements to support eCPR development. From this perspective we present the desirable features for development of eCPRs in the form of an incremental development model. This model builds using incremental steps of technological maturity delivering
required functionality to support the core functions of derivation and dissemination of eCPR knowledge required for the LHS.

In the following sections I have developed and described the incremental maturity model core components implemented in six steps to deliver a fully computable eCPR infrastructure that is consistent with the broader LHS (204). In addition to the incremental development model there is an important design, development and integration activity that needs to be considered across all stages of the maturity model. This activity includes the integration and re-design of clinical system user interfaces to incorporate the eCPR into the clinical workflow. The model therefore also considers workflow issues that are crucial across all steps of the model.

The purpose of this model is both conceptual and practical by:

- describing an incremental development path that promotes discussion and the collaborative formulation of a more ambitious technology enabled eCPR development vision
- providing the necessary functional roadmap to guide successful implementation of the aims of this research.

As a conceptual goal, it is hoped that those promoting use of eCPRs in their organisation can apply this model to their own technology environments to plan how to leverage their EHR data with a view to encouraging wider scale CPR uptake and use in the longer term.

4.2 A Working Definition of Interoperability for ICT Systems

As I discussed previously, any attempt to define a model for eCPRs needs to specifically address the technical barriers to implementation including interoperability. The ability to exchange data efficiently is a key requirement for the success and effective implementation of clinical systems and particularly CDSSs (205). The ability of ICT systems to effectively communicate and exchange data is known as interoperability (206). The interoperability of clinical systems has traditionally been problematic resulting in multiple systems with different representations of clinical information that are unable to communicate with each other.
It is useful on that basis to be specific regarding a working definition of interoperability that we can refer to during development of the model. Each level of the constructed eCPR model will identify the specific interoperability areas that it addresses. I am using the definition as provided by the Office of the National Coordinator for Health IT (207). This definition describes three interoperability layers:

- **syntax**- content and structure of information exchanged
- **semantics**-vocabulary/code sets/terminology that give the content a consistently defined semantic meaning through coding
- **transport and services**- the infrastructure components deployed and used to accomplish specific information exchange objectives and the actual delivery mechanism for information between parties.

### 4.2.1 Syntax

The starting point for achieving interoperability is the definition of agreed structured format for representing information that can be understood and processed by many different parties. In an effort to solve these problems a number of standards organisations have worked to develop accepted models of representing clinical information. In relation to the development of clinical ICT systems, interoperability may refer to the ability to process and the computable exchange of raw data (syntax or system interoperability).

Many different models have been proposed for representing clinical information (such as HL7, OpenEHR, CEN13606) to varying degrees of success (208-210). These typically employ and utilise an underlying data model. Some of these approaches provide a model that attempts to provide an abstract representation of real world clinical concepts (such as the HL7 v3 reference information model). Others take a more document oriented approach providing a model that represents documentation of real world clinical information (such as HL7 clinical document architecture or OpenEHR). Further development of standards is now emerging around the provision of web service driven architectures (136, 211, 212).
4.2.2 **Semantics**

The system interoperability proposed by these models may also be complemented by the use of binding of data items to clinical terminologies or classifications to provide for semantic interoperability. They define controlled vocabularies of clinical terms with associated unique codes and clinical meanings that can be used to unambiguously select and record clinical terms in any ICT application.

Clinical vocabularies and terminologies are an attempt to resolve these ambiguities providing defined clinical ‘codes’. But we have many different vocabularies and versions available that may have different philosophies about how to represent clinical data (e.g. Primary vs. Secondary care concepts).

Examples of vocabulary and terminology sets include:

- SNOMED (213)
- Read Codes (214)
- Unified Medical Language System (UMLS) (215)
- International Classification of Primary Care (ICPC2) (216)
- International Classification of Diseases (ICD) (217)
- LOINC (for lab tests) (218).

Some sample codes and the different formats used by each are demonstrated by concepts relating to Urinary Tract Infection: 68566005 (SNOMED, UMLS), U71 (ICPC2), K15..00 (Read Code), 599.0 (ICD 10), 53128-5 (bacteria in urine, LOINC). An example specifically in the primary care domain is the International Classification for Primary Care Version 2 (ICPC2)(216). Any developed model therefore also needs to consider the use of appropriate terminologies to provide for full interoperability. Vocabularies and terminologies are an important pre-requisite to facilitate structured data mining. Successful data aggregation and data analysis can only be done where we can guarantee that aggregated patient data is definitively talking about the same clinical concepts.

4.2.3 **Transport and Services**

Transport standards are considered separate to data representation and the actual meaning of that data. They describe networking standards support that
enables the actual transfer and exchange of that clinical information. Examples include Simple Object Access Protocol (SOAP) or Hypertext Transfer Protocol (HTTP). These represent low level network standards that will be considered outside the scope of this work low level transport standards can be used to build higher level software service level components that enable access to clinical services that support interoperability. The construction of a service-oriented architecture in the form of a ‘clinical evidence service’ in this research can be considered to be within the scope of the services interoperability definition.

4.3 Components of the eCPR Maturity Model
I have developed a multi-step maturity model for eCPRs as an important output of this research consisting of six incremental levels that define core functionality needed to support fully computable eCPRs (204). Each level provides functions that address particular interoperability needs. The six levels are shown in figure 4-1.

Figure 4-1 A Multi-step maturity model for the implementation eCPRs

4.3.1 Level 1: Literature Based CPRs
Interoperability Layers: not interoperable - standalone tool.
As previously discussed, the traditional format for distributing CPRs has largely been literature based with the associated limitations that result. This remains the obvious starting point for potential use of CPRs in clinical practice. A process of identification of CPRs from literature may identify potentially usable individual CPRs with respect to the particular clinical setting that they are to be used in. This may involve development of an electronic query based search strategy to identify candidate CPRs for further consideration (219). The identified literature based CPRs provide the starting point for development of subsequent eCPRs deployed as decision support tools.

As an example, the HRB Centre for Primary Care Research within the Royal College of Surgeons in Ireland used such an approach as the first stage in developing an electronic document based register of CPRs (86). A systematic review of published literature was conducted that identified CPRs specifically relevant to the family practice setting. Almost 800 published papers were identified up to 2010 indicating the increasing level of research and interest in this area. This approach is entirely dependent on literature and involves a large overhead in terms of on-going evidence curation, measured in terms of months of effort to do systematic reviews of the literature with associated staff costs. Work is on-going to review literature from 2010 onwards (219).

### 4.3.2 Level 2: Electronic Document Based eCPRs

**Interoperability Layers:** not interoperable - standalone tool.

An improvement on the traditional literature based CPR is to provide electronic document based equivalents. These eCPRs are not fully computable or integrated with other clinical systems in themselves but are documented as part of a collected register of rules in an electronic format with appropriate links to the on-line literature original document based sources. This is with a view to overcoming one of the initial difficulties in using CPRs by allowing more user friendly searching and identification of appropriate eCPRs for any presenting patient complaint. These search capabilities include searching on the lifecycle stage of the eCPR, the clinical domain or condition targeted by the rule, and the clinical settings in which it is suitable for deployment.
The HRB Centre for Primary Care Research within the Royal College of Surgeons in Ireland is developing an electronic register for CPRs in collaboration with the Cochrane Primary Health Care Field (220). This register has been developed from the systematic review of CPRs in family practice and contains 434 unique CPRs from almost 800 papers. The lack of validation and impact analysis of traditional CPRs was apparent where 238 (54.8%) have been validated at least once and only 12 (2.8%) have undergone impact analysis on either the process or outcome of clinical care (86). This register acts as a web accessible resource not only for researchers but also for clinicians at the point-of-care as shown in figure 4.2. This represents a valuable resource but is restricted by the lack of integration of the eCPRs with other electronic data sources such as the EHR.

![Electronic Register for Clinical Prediction Rules](image)

*Figure 4-2 The HRB Centre for Primary Care Research eCPR register*

**4.3.3 Level3: Electronic Computable Individual eCPR Tool**

**Interoperability Layers:** (syntax) - standalone tool, potential to integrate within a single organisation EHR.
The next level of the CPR maturity model involves development and deployment of specific literature based CPRs in the form of computable representations implemented in decision support tools in clinical practice. These computable rules implement an underlying model representation of the specific eCPRs being tested. The model representation may be specific to the individual rules used and may be embedded and specific to the particular information systems and settings in which they are deployed and tested. There may be some integration of the tools with EHR systems and clinical terminologies that are specific to the original development environment and associated patient populations (a narrow CPR validation in practice). As such they are tied to and may not be immediately accessible or generalisable beyond these specific applications or settings.

An important improvement though is that some limitations of traditional CPR lifecycle can be addressed through wider dissemination of eCPRs into clinical practice. The rule can be deployed electronically in a controlled clinical environment and made available to support subsequent validation and impact analysis efforts. This may take the form of randomised control trials testing the effectiveness of the electronic tool versus performance of a control group without access to it.

McGinn et al. demonstrated the potential impact of deploying eCPRs integrated with an EHR in family practice (12). This study implemented two well-validated individual CPRs as part of a randomised clinical trial in an electronic DSS that would ‘trigger’ when certain criteria were entered by the clinician into the EHR. Clinicians in the intervention group were invited to complete a risk score calculator and were given management recommendations based on this score. Unnecessary antibiotic prescribing for pneumonia patients was reduced by 9.2% and unnecessary streptococcal testing for pharyngitis patients was reduced by 12.1%.

The QRISK and QRISK2 tools for evaluating cardiovascular risk have been widely used as an electronic eCPR tool (126, 127). The QRISK2 score has been made available as an online stand-alone tool in its own right (221). In addition it has been more widely implemented in other EHR based systems.
such as the SystmOne EHR in the UK by making the underlying implementation available as open-source software along with software development kits to assist with developers that can be accessed from the QRISK2 website (221).

More generally, this approach demonstrates the benefit of using the EHR as a contextual trigger for eCPR decision support. Rather than having to search for the correct eCPR to use in any presenting patient case the EHR automatically provides the context for selection of appropriate eCPRs to consider.

4.3.4 Level 4: Service-oriented Generalised eCPR

**Interoperability Layers:** (syntax, services/transport) – open technical standards supporting accessibility from many different clinical applications within a single organisation; a separate underlying evidence knowledgebase connectable to many different end user applications and tools; lacks terminology integration so may limit access from other external organisational systems that use different clinical coding schemes.

The wider scale adoption of computable eCPRs beyond their initial development environments requires a more general model of electronic representation and deployment that is consistent with the wider vision of data consolidation and use as suggested by the LHS. As we have seen in the literature reviews, this suggests a service-oriented architecture providing access to web-based computable registers of eCPR resources that decouple provision and querying of eCPRs from the original deployment applications and clinical settings used in individual eCPR implementations. Computable eCPRs originally developed for use by individual hospital departments or family practices may then be made available as evidence to support wider scale dissemination, validation and impact analysis in tools developed for other associated departments, hospitals or family practices.

Service-oriented approaches to decision support delivery have become an increasingly important research topic (27, 135, 156, 222). The benefits of such an approach as specifically applied to decision support delivery include (223):
- implementation flexibility – the ability to develop decision support systems that can be accessed using multiple end-user devices such as desktop, web, tablet, or mobile applications
- centralised knowledge management - support for centralised maintenance of clinical knowledge that can then be rolled out to multiple applications or end-user devices
- support for multiple knowledge representations through a common web service interface – the underlying knowledge representation is hidden from end user applications as the service delivery is done through a common web service interface requiring no inherent of knowledge implementation
- separation of concerns – the decoupling of clinical applications from decision support web services provides for easier maintenance of applications as dependencies between them are reduced through the use of a standardised web service interface
- support for distributed decision support collaboration – a service oriented approach allows for the potential for decision support systems to be built in as modular reusable web services that can use each other to collectively provide decision support.

The provision of such a service is dependent on the implementation of a general model of representation for eCPRs that captures the core computable structures common to all eCPRs in general, and not just one particular eCPR specifically. In this way we can provide a general service that supports computable representations of potentially any eCPR rather than specific instances of them. The general model components that should be captured in the eCPR model include:

- a clinical outcome being tested by the eCPR
- the rule elements individually assessed as significant for consideration of the clinical outcome being tested
- a quantified weighting or scoring scheme attached to individual rule elements indicating their quantified significance to the clinical outcome
• the demographic context associated with the derivation population from which the scoring schemes were derived for use by an eCPR (this includes representation as appropriate of age, sex, and clinical setting related to any scoring scheme used derived for any rule, with potentially different demographic contexts and score schemes attached to a single rule to allow its application in different populations)

• a threshold based risk interpretation based on stratified score bands of the rule

• an optional clinical action or recommendation to be carried out in response to the interpreted risk based score bands.

The general model for eCPR representation provides a general structure to represent any individual eCPR in a computable way. In the literature review we highlighted the fact that CPRs can be considered as a form of discreet pieces of epidemiological knowledge. However there may potentially be many different CPRs relating to similar conditions or diagnostic outcomes. Appraising multiple CPRs for the same condition is typically done by systematic review of validation studies to compare rule performance (74). An example of such a review was carried out by Ceriani et al. providing a comparison and pooled analysis of studies that validated the performance of a number of rules for diagnosing pulmonary embolism (224). This provides for a comparative assessment of different rules along with the clinical settings in which they were used.

Different CPRs relating to the same diagnostic outcome may have scoring schemes that reflect different patient contexts depending on demographics or to different clinical settings. For example, it is possible that you may have two variations for male and female patients of a single CPR with the same diagnostic cues and diagnostic outcome. The scoring schemes could be different though to reflect the different relative weights of the diagnostic cues for each sex.

It is therefore also necessary to also implement a general diagnostic model of evidence which can be considered to be a general epidemiological model that represents all of the quantified evidence for each diagnostic outcome. Each eCPR can then be considered to be constructed from a specific subset of that
broader epidemiological evidence to be applied in a particular demographic context. eCPRs are therefore constructed from the general diagnostic model of evidence. In effect the general model of evidence ‘feeds’ a structured and contextualised ‘view’ of that evidence in the form of an eCPR.

4.3.5 Level 5: eCPRs with Terminology Services Integration

Interoperability Layers: (semantics) – semantically interoperable with many different ICT applications across multiple organisations through addition of standard clinical code bindings

The importance of integration of CDSS tools into the wider clinical workflow has been highlighted as a key factor for their broader acceptance and implementation success (93). The capability for binding of individual eCPR model terms with potentially many clinical terminologies and vocabularies to support wider semantic interoperability and broader uptake of eCPRs is crucial (225). This may be supported through provision of the service based eCPR models in conjunction with clinical terminology or vocabulary services that enable terminology lookup, binding and mapping of models to different vocabularies (226).

The integration of DSS tools with EHR systems based on coded patient data provides for the identification of workflow related patient events that can be used as a contextual trigger for initiating eCPRs as a form of decision support. In addition the patient record data itself can then be utilised to provide patient demographics or patient historical data that may be used to contextualise eCPR execution and selection of suitable scoring schemes based on the context of the particular patient under consideration. At a minimum then there is a requirement for the clinical evidence to be accessible to each EHR local site in an open and structured format that supports the clinical coding scheme required at the local site. The data quality, in terms of the how well and consistent coding of patient data has been actually been done, is a critical factor for this to work in practice, something that will be discussed in more detail in my conclusions.

The models of evidence should therefore allow for representation of clinical evidence independently of any given terminology or vocabulary. The explicit
representation of code binding concepts should be part of the models. This allows for association of ontology model terms to associated bindings of terminology codes. These terminology codes may be associated with eCPR initiating trigger events (in the form of a patient presenting complaint or diagnostic outcome under consideration) and with individual eCPR rule elements used for score calculation (provided by patient EHR data cues). Code bindings may be manually configured within the models themselves. However a more sustainable model would support the future integration of a fully functional vocabulary server (226). This would allow the eCPR models to do backend lookup and association of suitable target vocabulary terms, or actual runtime lookup of vocabulary mappings required for different coding schemes used in different EHR systems.

4.3.6 Level 6: Learning, Versionable eCPR

**Interoperability Layers:** (services) - Interoperable with many different ICT applications across multiple organisations; capable of deriving eCPRs.

The development and wider secondary use of aggregated sources of electronic patient data for research is a crucial part of the broader LHS vision. Whilst the LHS currently is still emerging and aspirational in nature there are already a number of existing aggregated sources of patient data to be found at organisational, local and national levels. These contain large amounts of longitudinal population health data that may be suitable for the application of data mining or statistical analysis techniques with a view to deriving actionable knowledge in the form of general evidence models that can then be used to construct specific eCPRs.

The literature review described the general data mining process and highlighted the suitability of structured data mining for clinical evidence discovery based on coded patient data. Data mining or statistical techniques can be applied to these repositories of patient data to detect and quantify diagnostic patterns found within the coded data. The associations discovered can then be transferred to the general model of evidence that was discussed in level 4 to give a broad epidemiological picture for any given diagnostic outcomes that have been data mined and analysed.
The QRISK score for cardiovascular risk assessment is an excellent example of an eCPR that was both derived and implemented electronically. It is notable in that it was electronically derived based on a retrospective analysis of an aggregated EHR primary care patient database rather than the traditional study based approaches previously described (126). It has also demonstrated the feasibility of implementing subsequent cyclic updates and revision of eCPRs as new clinical evidence emerges. This resulted in the derivation and implementation of the QRISK2 score (127). Annual revisions of the QRISK2 score are made available to take into account changes to discrimination and calibration as a result of underlying population changes.

At a local level in Ireland, the Irish College of General Practitioners (ICGP, the professional body for General Practice in Ireland), the HRB Centre for Primary Care Research and the National University of Ireland Galway (NUIG) are currently collaborating to create an Irish Primary Care Research Network (IPCRN) (182). The IPCRN allows for the extraction of anonymous patient data for audit and research purposes. 45 clinical practices now have the ability to audit their clinical data through it. The IPCRN can also facilitate the collection of national level patient data. Rich electronic sources of family practice data also exist in the United Kingdom (Clinical Practice Research Datalink (CPRD), The Health Improvement Network (THIN)), Canada (Canadian Primary Care Sentinel Surveillance Network) and the Netherlands (NIVEL) (178-181).
5 eCPR Framework Functional Design, Implementation and Results

This chapter describes the core implementation work done as part of this research. The conceptual framework developed in chapter 4 is applied in practice to produce a functional prototype implementation. This chapter is therefore the core of the research fieldwork. The work is broken down and described under the following broad sub-sections:

- 5.1. - explicit statement and rationale for key assumptions made relating to the implementation work done
- 5.2. - explicitly address and delineate the contribution of the author to the body of work done as part of the wider collaborative TRANSFoRm project
- 5.3 – 5.8 description of the implementation of the eCPR theoretical model that was outlined in chapter 4, focussing on higher maturity implementation descriptions of:
  - selection of clinical use cases
  - design of clinical models as part of level 4
  - construction of clinical models as part of level 4
  - construction of clinical evidence service to make clinical models accessible as a service as part of level 4
  - description of integration with terminology as part of level 5
  - description of a data mining approach supporting level 6 with case examples
- 5.9 - describe a diagnostic workflow within which the eCPR implementation can be accessed and used by third party tools
- 5.10 - evaluation of the work focussing on demonstration of its use by third party tools along with a conceptual evaluation of the decision support functionality it provides compared to best practice described in literature.
5.1 Assumptions and Scope Related to Implementation

Before describing the implementation of each of the levels of the model in detail, it is useful to document some specific assumptions I am making relating to the scope of the implementation work to be done. The following assumptions are being made:

- as previously discussed, there is recognition of the importance of assessing impact analysis of CPR based decision support tools on diagnostic accuracy and patient outcomes. This is a substantial piece of research in itself ideally requiring the conduct of a formal clinical trial in its own right and as such is considered outside the scope of this implementation work.
- backend evidence service models to support eCPR construction and dissemination are the focus of this work. Development of end user diagnostic decision support applications is not in scope but the use of the evidence service with diagnostic decision support tools will be demonstrated through integration with a separately developed third party decision support tool.
- developing aggregated sources of patient data for the purposes of data mining EHR data ourselves is not feasible so use is made of existing repositories of openly available clinically coded patient data, fully described in section 5.8, to demonstrate a data mining implementation.
- coded sources of electronic patient data will be used without an assessment of data quality of those sources, and an assumption is being made that the diagnostic labels are accurate and based on a recognised diagnostic gold standard in each case. There may be underlying limitations around identifying electronic cases based on the diagnostic label applied as it is not clear what was the gold standard used to apply the diagnostic label. There are significant limitations relating to data quality and bias of coded patient data that are considered and acknowledged in chapter 6.
- validation of derived data mined eCPRs would require application of the derived eCPRs along with an explicit gold standard for diagnosis to confirm a diagnosis. This is beyond the scope of this work and an
alternative approach will be to evaluate the derived eCPRs through comparison with best practice evidence based clinical guidelines obtained from clinical literature review.

- a formal validation of the diagnostic process itself is not proposed here as an assessment of the quality of the coded diagnostic labels applied in the TRANSHIS data cannot be done without knowing exactly how and when GPs coded the original data. Consideration is given to cases where there already is a leading diagnosis indicated in the coded patient data itself and then used for data mining.

5.2 TRANSFoRm Acknowledgement and Work Delineation

An LHS architecture based on the development of the eCPR model proposed in chapter 4 can now be described as summarised in figure 5-1.

![Figure 5-1 Architecture of an eCPR based learning health system](image)

This architecture consists of (30):

- a clinical evidence web service built on two underlying models of evidence as was described in chapter 4
- a data mining module to be applied to an aggregated source of EHR data in order to detect and empirically quantify diagnostic relationships
between presenting problems, recorded diagnostic cues, patient demographics and a recorded diagnosis

- a clinical evidence review tool to allow clinical review and curation of data mined evidence with a view to importing subsets of data mined evidence into the evidence service
- a provenance service that provides an audit function to track all activity of people, processes and datasets being used in the system (developed for TRANSFoRm but not described in this work)
- an eCPR construction tool to allow construction of eCPRs based on the imported and quantified data mined evidence
- decision support tool interface that consumes the eCPR evidence service to provide decision support linked to a patient EHR
- an evidence feedback loop to enable update of aggregated EHR data for the purposes of dynamically recalibrating the evidence base as new data is gathered and updated from the underlying patient population.

For the purposes of clarity it should be stated at the outset that the implementation of this architecture was done as part of my direct contribution to the TRANSFoRm project which was a collaborative effort. It is therefore good practice to clearly delineate what my contributions to each of the components were and clearly acknowledge contributions from any other individuals. My contribution to each component is described with appropriate references for each contribution where available.

The following components were primarily designed, developed and implemented by Derek Corrigan:

- framework for development of eCPRs as described in chapter 4 (204)
- the general evidence ontology model (227)
- the eCPR evidence ontology model (228)
- the eCPR Construction Tool.

The following components were designed, developed and implemented by Derek Corrigan with assistance from other staff as noted on implementation tasks:
• the clinical evidence service (implemented with Gary Munnelly) (229)
• integration of evidence service with decision support tool (implemented with Gary Munnelly and Samhar Mahmoud) (229).

The following components had secondary design input from Derek Corrigan as work package lead, with primary responsibility for design and implementation done by other members of the TRANSFoRm project:

• definition of a data mining methodology (172)
• implementation of a clinical evidence review and curation tool
• development of provenance tracking for decision support
• development of a decision support tool interface supporting EHR integration (229).

5.3 Level 1 - Selection of Working Clinical Use Cases

In order to develop and test the clinical evidence models it is useful to consider development of a specific clinical scenario or use case. The literature review described the diagnostic process in primary care. A clinical scenario should assist with working through the distinct diagnostic steps identified as part of that diagnostic process. At a minimum the clinical scenario should describe:

• a common presenting patient complaint in primary care practice with an associated demographic context
• a selected and limited number of differential diagnoses to consider relating to that patient complaint
• the general diagnostic evidence as described in literature review and clinical guidelines to support or rule out these selected diagnoses
• a selected number of well documented clinical prediction rules relating to some of these diagnoses for which we wish to develop computable eCPRs.

On this basis I have selected the case of a female patient presenting with a presenting complaint (or reason for encounter) of: ‘abdominal pain’ (230).

A large number of potential diagnoses may need to be considered by a clinician when encountering this broad clinical complaint. For our scenario there are a
number of differential diagnoses that might potentially be considered. I will consider a particular of subset of differentials that have been specifically selected for our purposes that collectively demonstrate the diagnostic characteristics that were previously discussed as being potentially problematic for clinicians (231). The selected examples that I will consider are:

- ectopic pregnancy
- pyelonephritis
- urinary tract infection
- Crohn's disease
- appendicitis
- bowel cancer
- irritable bowel syndrome
- bacterial enteritis (230).

There are two potential options for defining and obtaining data to support such a clinical use case. These are:

- systematic review of clinical literature for diagnostic evidence and associated CPRs to support the stated clinical use case
- obtain electronic coded primary care that can be data mined with a view to populating models of evidence data to support the stated clinical use case.

Initial clinical evidence models will be manually constructed and populated with diagnostic clinical evidence to enable model testing. I will subsequently discuss an example of population of the models using data mined evidence in 5.9. Diagnostic evidence gathering to support each selected differential diagnosis was done by carrying out a systematic review of the literature for each diagnostic condition. Refer to appendix A for an example of the of the search strategy employed for these reviews. Appendix K shows an example of the output and format of such a review of diagnostic evidence, in this case for urinary tract infection.
5.3.1 The Alvarado Score as a CPR Example

To illustrate development of our eCPR model, I am implementing a computable version of a well described CPR in literature for diagnosing appendicitis called the Alvarado Score (232). This is an example of a well studied CPR which we will use as a clinical example to illustrate development and validation of the initial eCPR model based on an existing literature derived CPR. The rule illustrates the core structured concepts that define a CPR; a set of diagnostic cues, a set of cue associated scores, a stratified risk based interpretation of the total rule score and a set of clinical decisions associated with each defined risk category (232).

This rule categorises the risk of patients having potential acute appendicitis based on the presence or absence of 8 diagnostic indicators. The risk of appendicitis is expressed as three score based risk categories with associated recommended treatment options. This rule has been designed to be suitable for use in primary care and is based on the presence of diagnostic cues without the need for imaging (232). A summary of the Alvarado score is shown in figure 5-2.
Reviews have highlighted the importance of capturing the epidemiological characteristics or demographic context of the derivation study population. Clinical performance of the Alvarado score has been shown to vary in different populations depending on gender and age, performing best for adult males (91) (232). The Alvarado score showed a relative risk ratio close to 1 (the ratio of predicted number of cases to the actual number of cases) for all three risk categories for men: low risk 1.06, intermediate risk 1.09 and high risk 1.02 (91). The relative risk ratios show over-prediction across all three risk categories for women however: low risk 5.35, intermediate risk 1.82 and high risk 1.14 (91).
For the purposes of data-mining and constructing an actual eCPR the clinical case of Urinary Tract Infection will be considered as it is more frequent in women and will be discussed later. This demographic context therefore should be reflected in any model design so that this variability can be captured.

5.4 Level 4: Generalised Models of Clinical Evidence

Delivery of level 4 functionality of the maturity model requires:

- construction and population of a general evidence ontology model (229)
- construction and population of general eCPR ontology model (228)
- making the models available and accessible to third party applications as web service (204).

5.4.1 Ontology Derivation and Validation Based on Competency Questions

An ontology is used as the model to represent the diagnostic cues that provide the evidence supporting diagnosis of a number of presenting clinical conditions. A number of different methodologies are described for designing and implementing ontologies are found in literature (233). The methodology I have used is to employ an ‘application focussed’ ontology design. This considers design of the ontology within the context of the functional requirements of the actual application that will use the ontology (rather than designing the ontology in isolation for its own sake). Using this approach the functional requirements of the application can be expressed using ‘clinical competency questions’. These functional requirements provide the starting point for the design, implementation and validation of the required ontologies. The ontology design methodology I have used is based on the design practices advocated by the work of Gruninger and Fox (234). This methodology advocates the following steps to assist with ontology definition and construction:

- draw up a list of informal questions that correspond with functional requirements of the application that your proposed ontology should be able to answer
• deconstruct those questions by identifying formal domain concepts in the form of knowledge domain objects, relationships and domain instances that are required to answer the competency questions
• construct the ontology based on the formal concepts you have identified
• express the informal competency questions using formal ontology query languages and your constructed ontology
• validate your ontology through clinical use cases to demonstrate that the results generated from formal ontology questions correctly answer your clinical use cases.

5.4.2 Definition of Functional Requirements as Clinical Competency Questions

In considering the model design requirements it is necessary to first consider the functional requirements of any application that will use those developed models. Based on our previous description of the diagnostic process the requirements for these models are to support a diagnostic process that is capable of suggesting potential differential diagnoses to consider for a particular clinical problem, supported by associated eCPRs to rule out the most unlikely ones to consider. This is done based on the specific patient case data that is presented as part of the clinical consultation with the GP, and could be manually entered or extracted from a patient EHR and passed to the models hosted by the clinical evidence service.

I have defined our model requirements based on the different diagnostic questions it needs to be able to answer of both the general evidence model and the eCPR evidence model. The functional requirements can be stated as clinical questions we wish to be able to ask of our models. Each of these questions can then be examined to identify potential ontology concepts that we will need to model.

A particular clinical example can be used to frame the competency questions with a view to identifying the general clinical concepts that should be included in the developed model. It is not necessary to develop competency questions for all of the clinical conditions from our use cases described previously in section
5.3. What is important is that all of the clinical questions that need to be answered by our model are described for the chosen clinical example selected.

Using the example of a patient presenting with ‘dysuria’ and suspected ‘Urinary Tract Infection’ we can identify the questions shown in table 5-1 as functional requirements that we would want to be able to answer using the finished workbench tool, with potential ontology concepts to be modelled highlighted in red:
Table 5-1 Clinical competency questions to describe requirements for developing the ontology models

<table>
<thead>
<tr>
<th>No</th>
<th>Clinical Competency Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>What are the <em>differential diagnoses</em> to consider for a <em>patient</em> presenting with a <em>reason for encounter</em> (RFE) of abdominal pain?</td>
</tr>
<tr>
<td>2</td>
<td>What are the <em>diagnostic symptoms</em> that support a <em>diagnosis</em> of Urinary Tract Infection?</td>
</tr>
<tr>
<td>3</td>
<td>What are the <em>diagnostic signs</em> that support a <em>diagnosis</em> of Urinary Tract Infection?</td>
</tr>
<tr>
<td>4</td>
<td>What are the <em>diagnostic risk factors</em> that support a <em>diagnosis</em> of Urinary Tract Infection?</td>
</tr>
<tr>
<td>5</td>
<td>What are the <em>diagnostic tests</em> that may be carried out to support a <em>diagnosis</em> of Urinary Tract Infection?</td>
</tr>
<tr>
<td>6</td>
<td>How strong is the <em>quantified association</em> between a positive <em>diagnosis</em> of Urinary Tract Infection and the presence of the <em>diagnostic cue</em> of dysuria (pain on urination)?</td>
</tr>
<tr>
<td>7</td>
<td>In what <em>demographic context</em> and <em>clinical setting</em> was a <em>quantified association</em> between a <em>diagnosis</em> and a <em>diagnostic cue</em> derived?</td>
</tr>
<tr>
<td>8</td>
<td>What are the <em>CPRs</em> associated with the <em>differential diagnosis</em> of appendicitis?</td>
</tr>
<tr>
<td>9</td>
<td>What are the <em>cues, criteria</em> and associated <em>scores</em> of the Alvarado score?</td>
</tr>
<tr>
<td>10</td>
<td>What are the <em>scoring interpretation schemes</em> of the Alvarado score?</td>
</tr>
<tr>
<td>11</td>
<td>What are the <em>population characteristics</em> associated for application of the Alvarado score?</td>
</tr>
<tr>
<td>12</td>
<td>What is the <em>clinical setting</em> associated for application of the Alvarado score?</td>
</tr>
<tr>
<td>13</td>
<td>What are the <em>supporting literature</em> sources for the Alvarado score?</td>
</tr>
<tr>
<td>14</td>
<td>What is the current <em>version number</em> and <em>CPR maturity level</em> of the Alvarado score?</td>
</tr>
</tbody>
</table>
5.4.3 **General Model of Evidence Concepts**

The ontological concepts I have highlighted in our clinical competency questions are crucial to the representation of clinical evidence in a manner that is fit for implementation as part of a broader diagnostic strategy. We previously discussed how CPRs can be deployed as part of a wider diagnostic strategy particularly around ‘ruling out’ proposed differential diagnoses for a particular problem. The following ontology characteristics are necessary to support such a diagnostic strategy.

5.4.3.1 **RFE - Ontology Concept**

A single RFE that expresses the primary reason or problem that led the patient to consult with the clinician. In a situation where the patient is suffering from more than one clinical problem, the patient could present with potentially more than one reason for encounter. In the case of our clinical scenario the patient presents with a single RFE of ‘Abdominal Pain’.

5.4.3.2 **Diagnosis - Ontology Concept**

A diagnostic outcome or label for which diagnostic evidence can be provided to either rule in or rule out that diagnosis. A diagnosis can be a differential diagnosis of more than one particular presenting patient problem (RFE).

5.4.3.3 **Cue - Ontology Concept**

Each Diagnosis will have associated diagnostic Cues. For example ‘Urinary Tract Infection’ has a symptom called ‘Dysuria’ e.g. ‘Dysuria’ is a clinical symptom reported by the patient. This concept also includes a number of sub concepts (Sign, Symptom, Test, Risk) that represent four underlying areas of diagnostic investigation relating to any reported patient complaint. These are:

- patient reported symptoms
- clinician observed and recorded signs
- patient risk factors including population demographics, genetic disposition, lifestyle factors or previous historical conditions
- clinical tests that are associated as clinical evidence to support a diagnosis.
5.4.3.4 Context - Ontology Concept
I have previously stressed the importance of identifying the demographic context that is suitable for the application of any piece of clinical evidence. This is largely defined by the original derivation population from which the diagnostic evidence was derived from. This will inform the characteristics of the population that the clinical evidence is most suitable for application to. These concepts describe key demographic and clinical setting contexts that are associated with the with a particular CPR such as age, sex, ethnicity, clinical setting, country.

   e.g. 'Male', 'Female', 'Caucasian', 'Primary Care', 'Ireland'.

5.4.3.5 Quantification - Ontology Concept
This is a key concept to provide for the representation of data mined clinical evidence. This concept represents a quantified measure of association providing an empirical measurement of how closely associated a particular positive diagnosis is based on the presence of a particular patient problem (RFE), or how closely associated a positive diagnosis is based on the presence of a particular diagnostic cue.

These quantifications can be in the form of a positive or negative likelihood ratio for example which determines how much more likely a positive diagnosis is in the presence of a particular piece of diagnostic evidence. Each quantification exists within a particular demographic context defined by the underlying population from which it was calculated. The relative strength of the association may therefore only be valid when considered as evidence for a patient of the same demographic context from which the quantification value was originally derived.

More than one quantification for different demographic contexts may exist for the same RFE/Cue/Diagnosis relationship. For example the quantified positive likelihood ratio of ‘dysuria’ as an indication of ‘Urinary Tract Infection’ could vary in the demographic context of being found present in males and females.

5.4.3.6 Define Ontology Relationships
The ontology models support a number of named ontological relationships that are bidirectional in nature and define the relationships needed between different ontology concepts. This means that either a ‘top-down’ approach, starting from
a reason for encounter working down to individual diagnostic cues, or a ‘bottom-up’ approach starting from a diagnostic cue and working up to associated diagnoses can be employed. Each relationship has corresponding inverse relationships so as triples we can represent diagnostic evidence using directional statements in the form:

- ‘Urinary Tract Infection’ - ‘hasSymptom’ - ‘Dysuria’
- ‘Dysuria’ - ‘isSymptomOf’ - ‘Urinary Tract Infection’

The clinical competency questions are then deconstructed into ontology concepts along with the required relationships (defined in both directions) that are needed to link them. The clinical concepts to support a general model of diagnostic evidence can be modelled as an ontology of clinical evidence. The initial proposed ontology model is described in terms of the concepts and relationships identified and shown in figure 5-3. The ontology diagrams show relations in one direction only but definition of inverse relationships are also implied. This allows representation of the relationships between a presenting patient reason for encounter and the associated candidate differential diagnoses to consider. The evidence relating to any particular diagnosis is captured as associated diagnostic cues, of which there are cue sub-concepts to represent clinician observed signs, patient reported symptoms, risk factors and clinical tests.
5.4.4 **eCPR Model of Evidence Concepts**

This provides a model implemented as an ontology of eCPRs that can be used to deploy eCPRs in a computable format that is accessible using open standards by third party tools via a web service (227, 235). The ability of such a model to capture the demographic context associated with scoring schemes implemented for an eCPR is an important requirement. This allows for potentially more than one threshold based scoring scheme relating to different population contexts that can be associated with any individual eCPR. Two different scoring schemes for the same eCPR might exist for male and female sexes for example, or different scoring schemes could exist for populations from different countries. This can address the static limitations of traditional literature based CPRs by providing for future development of dynamic population sensitive eCPRs that have scores that can be changed or extended beyond the demographic context of the original derivation population, laying the groundwork...
for more evolved eCPRs derived via data mining techniques. The model is shown in figure 5-4.

![CPR ontology model concepts and relationships](image)

**Figure 5-4 CPR ontology model concepts and relationships**

### 5.4.4.1 CPR – Ontology Concept
This represents a named eCPR associated with a particular diagnosis that has been obtained from literature or derived using data mining. For example ‘Appendicitis’ will have an associated ‘Alvarado Score’ CPR. The eCPR has associated property values that capture the version number of the eCPR, and the CPR lifecycle maturity (level 1 to level 4) as was explained in section 2.6.6. This can be used to select eCPRs that are considered ‘good quality’ and have been validated and impact assessed in clinical practice.

### 5.4.4.2 CPRElement – Ontology Concept
An eCPR is constructed from a number of eCPR elements. Each eCPR element defines a diagnostic cue that is scored as part of the eCPR along with the particular diagnostic criteria to be checked against it.
5.4.4.3 **Cue – Ontology Concept**

A diagnostic cue that identifies a significant piece of clinical evidence for consideration related to the outcome of applying a particular eCPR.

5.4.4.4 **Criteria – Ontology Concept**

The criteria to be checked for a particular diagnostic cue associated with the eCPR. The criteria could be simply checking the presence or absence of the cue or it may involve comparison against a particular testing value or range of values. Where the criteria is satisfied we also capture what the user interpretable whole number score for satisfying the diagnostic criteria is in the context of that particular eCPR. The score is in effect also a quantification and exists within a particular demographic context defined by the underlying population from which it was calculated.

5.4.4.5 **CPRScoreScheme – Ontology Concept**

The total score for the eCPR when applied in the context of a particular patient case needs to be clinically interpreted as a range of scores defining risk bands. These risk bands represent the threshold approach to risk interpretation as previously discussed. The score scheme provides the bands, risk categories and associated clinical actions to be carried out for each risk category.

5.5 **Ontology Model Construction - Development Tools**

Protégé is a well established tool for construction of ontologies. After reviewing available tools it was considered as suitable choice for the development of the general model of evidence ontology and the eCPR ontology models (236). Protégé version 4.3 has been used for the ontology modelling work. The ontology concepts previously identified were implemented along with associated relationships and instances of each class defined by the clinical content required for the patient safety use cases.

The steps required for construction of the ontology were:

- defining the ontology class structure (figure 5-5)
- defining object relationships existing between classes (figure 5-6)
- defining instances of classes to represent our clinical knowledge related to our specific patient safety use cases (figure 5-7)
Protégé uses web ontology language (OWL) and resource description framework (RDF) standards to represent the constructed ontology as a text file that can then be loaded and hosted in an ontology database called a triple store (150, 152).
Figure 5-6 – Defining ontology relationships in Protégé
5.5.1 Modelling Multi-morbidity

In the introduction I have already made reference to the problems and limitations of paper-based guidelines in addressing patients with multi-morbidities. The developed model must therefore consider the possibility of a patient presenting with more than one condition at the same time. A number of distinct models of morbidity can be implemented (237). They are represented pictorially in figure 5-8 where circles represent a diagnosis and squares represent a diagnostic cue. Suggested models to consider are:

- **a - associated liabilities model** establishes dependent relations between individual diagnostic cues
- **b - multiformity model** establishes relations between diagnostic cues and diagnoses with a cue being associated with potentially many (perhaps multi-morbid) conditions
- **c - causal model** – relations between diagnoses
• **d - independence model** – multi-morbid diagnoses modelled as separate individual entities

• **e - hybrid model** – mix of associated liabilities and multiformity models.

![Figure 5-8 Models of morbidity](image)

For the purposes of representing clinical evidence found in both literature and from data mining, the multiformity model option (b) was most suitable and has been implemented as part of the general model of clinical evidence. The same individual diagnostic cues may have associated (and quantified) relationships with more than one diagnosis. For example the diagnostic cue ‘dysuria’ may be related to more than one diagnosis ‘urinary tract infection’ or ‘pyelonephritis’.

### 5.5.2 Evidence Service Construction and Architecture

The Protégé tool allows us to define and populate our required diagnostic model structures. The output from Protégé is simply a text file definition of the model. In order to provide a platform for hosting our models that allows for multi-user access as a clinical evidence web service, a defined application architecture is required. This architecture will also support subsequent development of data mining tools to dynamically generate evidence from electronic sources and update the models with the discovered evidence through applications rather than manual update.

The Protégé ontology construction tool as previously described allows definition of the model structure but does not provide a platform to make it available as a
multi-user application. Ontology development has matured and addressed the issue of providing scalable multi-user access to content through the development of what are called ‘ontology triple stores’, the equivalent of a database for hosting ontologies. These provide an implementation platform to host developed ontologies that is suitable for access by multiple users and can be programmatically manipulated to update and query content as necessary.

We have selected a well regarded triple store developed by the OpenRDF foundation known as Sesame (238). This triple store can provide for query formulation, testing and future dynamic programmatic update of ontology content.

The clinical evidence service consists of three implementation layers that support this:

- the persistence layer that provides multi user access and dynamic content update through a Sesame data triple store to host the protégé ontology upon which the evidence service is constructed (149, 238)
- a service layer providing a fully functional REST based web based interface with defined query methods to access questions of the ontology content from any third party tool (239)
- a client layer provided as a client side library used by the TRANSFoRm decision support interface to provide for exchange of patient data from the CDSS to the evidence service and return of evidence recommendations to the CDSS interface (240).

The implementation technologies for each of the three implementation layers are summarised in figure 5-9 and will be described in detail later.
Figure 5-9 Architecture implementation technologies

The fully implemented architecture and components are shown in figure 5-10 and described in detail in the subsequent sections.
Figure 5-10 Clinical evidence architecture overview
5.5.3 Ontology Triple Store Selection and Performance Considerations

Sesame provides a rich persistence platform around which a robust evidence service can be developed and hosted. A number of triple stores were initially considered but Sesame was selected based on its open source support and its favourable performance characteristics when compared to other available platforms (238, 241).

Sesame provides for hosting of an ontology using a traditional underlying database such as Java Native Database or MySQL as a datastore. The Sesame platform allows for the browsing of ontology through a web based workbench that allows access to any of the ontology concepts and content defined in the clinical evidence ontology as shown in figure 5-11. Results are shown in the form of ‘subject’, ‘predicate’, ‘object’ triples.

![Sesame triple store workbench browser](image)

Figure 5-11 Sesame triple store workbench browser

5.5.4 Development of SPARQL Queries

Sesame allows querying of the content using structured ontology querying languages such as SPARQL Protocol and RDF Query Language (SPARQL)
which is a triple store equivalent of database SQL queries (shown in figure 5-12) (149). Using the named ontology relationships that link concepts, these flexible queries allow implementation of ‘top-down’ or ‘bottom-up’ diagnostic strategies.

For example we might wish to do a simple query to select the symptoms of Urinary Tract Infection. Because ontology relationships are directional in nature we can either use the relationship that links a diagnosis to a symptom (the ‘hasSymptom’ relationship) or use the relationship that links a symptom to a diagnosis (the ‘isSymptomOf’ relationship). The unknown symptom variables in the query are represented with a ‘?’ as ‘?anySymptom’.

A top down query from a diagnosis to give all symptoms is:

‘SELECT ?anySymptom WHERE (UrinaryTractInfection hasSymptom ?anySymptom)’

A bottom up query from all symptoms of a diagnosis is:

‘SELECT ?anySymptom WHERE (?anySymptom isSymptomOf UrinaryTractInfection)’
As we have seen, Sesame allows a user to interact directly with it using a set of interactive user tools. The Sesame architecture also allows externally developed software applications, such as data mining tools, to access and update data contained in the Sesame triple store just like a database. Sesame defines a set of programmable interfaces and methods as an open source application programming interface (API) allowing it to be used as an underlying data store for other developed applications. The applications can connect to Sesame and execute SPARQL queries to retrieve or update data as required. This provides the building blocks for development of the clinical evidence service that sits on top of the application triple store. For a complete example of code that uses the Sesame API to execute a SPARQL query refer to Appendix B.
5.5.6 **Ontology Model Validation**

Before developing the clinical evidence service itself it is necessary to consider the validation of our ontology model structure. In sections 5.4.1 – 5.4.4 we described the design of our ontology requirements in the form of clinical competency questions. If our ontology is designed correctly we should be able to express all our competency questions as formal ontology queries that generate correct clinical results with respect to our selected clinical use cases when executed. As an example we can consider the case of representation of the Alvarado Score for appendicitis. Previously we saw that competency questions were deconstructed to identify the required formal ontology concepts and defined relationships that exist between them.

Each concept is a general concept that does not describe specifics for a particular clinical case. Within that concept we can define instances of those concepts that are required to be populated into the ontology to support our particular clinical use case. The concepts are defined to be general enough to support other clinical use cases above those that we have selected here.

For testing I will consider a patient presenting with ‘abdominal pain’ being assessed for a diagnosis of ‘appendicitis’ using the Alvarado Score. Examples of the concepts and concept (or class) instances and relationships required to electronically represent these are described in table 5-2.
Table 5-2 Summary of CPR ontology concepts and instances

<table>
<thead>
<tr>
<th>Class Name and Description</th>
<th>ClassInstance / Relationships</th>
</tr>
</thead>
<tbody>
<tr>
<td>EvidenceRFE - The patient reported reason for encounter (RFE)</td>
<td>AbdominalPainRFE</td>
</tr>
<tr>
<td>evidenceDifferentialDiagnosis</td>
<td></td>
</tr>
<tr>
<td>EvidenceDiagnosis - A differential diagnosis of a particular RFE</td>
<td>Appendicitis</td>
</tr>
<tr>
<td>ClinicalPredictionRule - A versioned CPR associated with a particular diagnosis with links to supporting literature URLs</td>
<td>AlvaradoScore1_0</td>
</tr>
<tr>
<td>hasCPR</td>
<td></td>
</tr>
<tr>
<td>ClinicalPredictionRuleElement - One individual element of the CPR that is associated with one cue and the criteria to apply to it</td>
<td>AlvaradoScoreElement1</td>
</tr>
<tr>
<td>EvidenceCue - An associated sign, symptom, risk or clinical test</td>
<td>ReboundTenderness</td>
</tr>
<tr>
<td>EvidenceCriteria - The criteria and weighted rule score associated with a ClinicalPredictionRuleElement where the criteria is true</td>
<td>isPresent</td>
</tr>
<tr>
<td>hasScoreInterpretation 1</td>
<td></td>
</tr>
<tr>
<td>ClinicalPredictionRuleScore - A score range to be used for clinical interpretation of the rule along with the textual interpretation of that score level</td>
<td>AlvaradoScoreLevel3</td>
</tr>
<tr>
<td>hasStartScore 7 hasEndScore 10 hasScoreInterpretation “Surgery”</td>
<td></td>
</tr>
<tr>
<td>EvidenceContext - A group of classes that defines the evidence population demographics used to derive the rule</td>
<td>Adult, Male, Ireland</td>
</tr>
<tr>
<td>EvidenceClinicalEnvironment - The clinical setting or context in which the rule was derived and is suitable for application</td>
<td>PrimaryCare</td>
</tr>
</tbody>
</table>

5.5.7 Model Testing

The initial testing and validation of our model is done with respect to the original Alvarado Score definition as obtained from clinical literature sources and described in section 5.3.1. If our construction has been done correctly we should be able to express our original clinical competency questions in 5.4.2 and functional requirements as formal ontology queries using our chosen ontology query language (SPARQL). The results returned from these queries should be consistent with respect to the literature definition of the Alvarado Score. Using the Sesame infrastructure and the implemented ontology model we have developed formal SPARQL queries to answer the clinical competency questions for our clinical use case as shown in table 5-3.
The query results returned from the SPARQL queries are consistent with the representation of the Alvarado Score as described in literature demonstrating the conceptual feasibility of computable ontology based CPRs. Flexible queries can answer clinical questions required of computable CPRs.

### Table 5-3 Clinical competency questions implemented as formal ontology queries with results

<table>
<thead>
<tr>
<th>SPARQL (Protocol and RDF Query Language)</th>
<th>Query Result (Instance Relation Value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELECT ?anyDifferentialDiagnosis WHERE {?anyDifferentialDiagnosis isDifferentialDiagnosisOf AbdominalPainRF}.</td>
<td>Appendicitis, BacterialEnteritis, ChronicDisease, CorPulmonale, EctopicPregnancy, Pyelonephritis, UrinaryTractInfection</td>
</tr>
<tr>
<td>SELECT ?anyCPR WHERE {?anyCPR isCprOf Appendicitis.}</td>
<td>AlvaradoScore1.0</td>
</tr>
<tr>
<td>SELECT ?anyCueElement ?anyProperty ?anyValue WHERE {?anyRuleElement isRuleElementOf AlvaradoScore1.0. ?anyCriteriaElement isCriteriaOf ?anyRuleElement. ?anyCueElement isCueElementOf ?anyRuleElement. ?anyCriteriaElement ?anyProperty ?anyValue. ?anyProperty rdf:type owl:DatatypeProperty.} ORDER By ?anyCriteriaElement</td>
<td>MigrationOfPain isPresent true, MigrationOfPain hasScoreInterpretation 1, Anorexia isPresent true, Anorexia hasScoreInterpretation 1, Nausea isPresent true, Nausea hasScoreInterpretation 1, RightLowerQuadrantTenderness isPresent true, RightLowerQuadrantTenderness hasScoreInterpretation 2, ReboundPain isPresent true, ReboundPain hasScoreInterpretation 1, ElevatedTemperature isPresent true, ElevatedTemperature hasScoreInterpretation 1, Leucocytosis isPresent true, Leucocytosis hasScoreInterpretation 1, Leucocytosis hasScoreInterpretation 2, WhiteBloodCellShiftLeft isPresent true, WhiteBloodCellShiftLeft hasScoreInterpretation 1</td>
</tr>
<tr>
<td>SELECT ?anyScoreElement ?anyProperty ?anyValue WHERE {?anyScoreElement isScoreSchemeOf AlvaradoScore1.0. ?anyScoreElement ?anyProperty ?anyValue. ?anyProperty rdf:type owl:DatatypeProperty.} ORDER By ?anyScoreElement</td>
<td>AlvaradoLevel1 hasScoreInterpretation &quot;Discharge&quot; AlvaradoLevel1 hasStartScore1 AlvaradoLevel1 hasEndScore4 AlvaradoLevel2 hasScoreInterpretation &quot;Observation/Admission&quot; AlvaradoLevel2 hasStartScore5 AlvaradoLevel2 hasEndScore6 AlvaradoLevel3 hasScoreInterpretation &quot;Surgery&quot; AlvaradoLevel3 hasStartScore7 AlvaradoLevel3 hasEndScore10</td>
</tr>
<tr>
<td>SELECT ?anyLiteratureURL WHERE {evd:AlvaradoScore1.0 evd:hasLiteratureURL ?anyLiteratureURL.}</td>
<td><a href="http://www.biomedcentral.com/content/pdf/1741-7015-9-139.pdf">http://www.biomedcentral.com/content/pdf/1741-7015-9-139.pdf</a></td>
</tr>
</tbody>
</table>
5.6 Level 4: Service Oriented Provision of Evidence Content

The ability to disconnect the decision support knowledgebase and make it available as a web accessible service was highlighted as required functionality for level 4 of our eCPR maturity model. I now describe the implementation of such a service to support provision of a clinical evidence service that is independent of any specific decision support application.

5.6.1 Jersey Java implementation of REST Services

The service layer of our implementation provides a fully functional Jersey REST based web interface with defined REST methods (or endpoints) to allow parameterised querying of diagnostic questions based on patient data supplied from a third party consumer (239). In order to provide a fully functional evidence service without the need for underlying knowledge of SPARQL or the ontology structure, we have implemented a broad range of parameterised REST based methods that can be used to run standard diagnostic queries against the evidence server. These methods are accessible as structured web based URLs that can return XML, JSON or RDF formats to any third party tool that wishes to use it formats (150, 242, 243). Appendix B and C provides an example of using Jersey Java libraries and the Sesame API to execute a SPARQL query through a REST endpoint that defines a structured URL path to access this functionality and return XML results to the calling application.

5.6.2 Clinical Evidence Service Interfaces and Methods

Using structured evidence service endpoints we can access any ontology content with results returned as XML (default), JSON or RDF. The REST query to access the differentials to consider for a patient presenting with abdominal pain for example is:

http://phaedrus.scss.tcd.ie/munnellg/ClinicalEvidenceRESTService/interfaces/query/rfes/differentials/AbdominalPainRFE

The output from the query is shown in figure 5-13 (note code bindings are hidden and will be discussed in the next section).
Figure 5-13 An evidence service reply describing differential diagnoses to consider related to a presenting reason for encounter of abdominal pain

To access the cues supporting diagnosis of urinary tract infection (output shown in figure 5-14 with code bindings hidden) the query is:

http://phaedrus.scss.tcd.ie/munnellg/ClinicalEvidenceRESTService/interfaces/query/differentials/cues/UrinaryTractInfection
Figure 5-14 An evidence service reply describing a symptom collection for Urinary Tract Infection.

All the currently defined REST endpoints are available from the evidence server by accessing a server root URL and adding the appropriate REST path from the tables shown in Appendix D.

5.6.3 Sesame REST API Support for Ad-hoc Queries

The clinical evidence service implements fixed web service REST methods that implement standard basic queries necessary for querying the ontology content.
Sesame provides flexibility beyond this by providing standard functionality to submit SPARQL queries directly against the server through its own web accessible interface. With knowledge of the clinical evidence ontology structure this allows for any third party application to formulate and submit SPARQL queries for the purposes of creating their own ad-hoc querying directly against the server returning results in RDF or XML formats. This can provide for more advanced queries beyond those provided though the defined clinical evidence REST service.

5.7 Level 5: CPRs with Terminology Services Integration

The previously discussed implementation of the eCPR maturity model levels up to and including level 4 has provided diagnostic models that are made available using an openly standards in the form of an openly accessible web service. This provides the basis of support for system interoperability to third party application consumers.

The correct semantic interpretation of the information provided by the web services requires that we address semantic interoperability concepts. Semantic interoperability support is provided for by extending our ontology models with the addition of a ‘code binding’ concept that is separate and independent of the clinical concepts. This is used to associate potentially many different clinical terminology codes for any single ontology RFE, cue or diagnosis as shown in figure 5-15.

Coding schemes are not always defined in a very ‘clean’ way. In many coding schemes there may be more than one code that could describe a clinical concept. Where multiple synonym codes exist for an ontology concept, the code binding concept has an ‘isPrimary’ attribute that can be set to ‘true’ or ‘false’ and indicates which code binding is considered the primary code binding and textual description. In this way the evidence can be triggered by any of a number of suitable codes but the text description shown in any target consumer application can be set to the most appropriate code available.
Localisation support to allow easier searching by a third party consumer for ontology terms using locally defined synonyms is provided for by the ‘synonym’ concept. Local synonyms are not necessary for any implementation but do allow for a more user friendly experience by facilitating the use of local terms while still making use of and underlying structured coding scheme. A core coding scheme may be sufficient without local synonyms. Local synonyms need to be manually associated with coding scheme terms so there is an overhead of effort required to maintain them that depends on the number of synonyms needed.

An example of concepts and associated instances (in red) of the diagnostic cue concept for a patient history of irritable bowel syndrome, with an associated NHS read code ‘14CF.00’ and local synonym ‘HO IBS’ is shown in figure 5-16. The ontology can support other coding schemes including ICPC2, ICD10, SNOMED and UMLS (213, 215-217).

For the purposes of testing integration of the developed web service with a third party EHR I have specifically supported interoperability with a single EHR
vendor called ‘In Practice Systems’ in the UK who develop the Vision EHR that makes use of NHS read codes version 3 (214).

Figure 5-16 Example of cue ontology concept instance for ‘history of irritable bowel syndrome’ with associated code and synonym.

The evidence service also makes code bindings and synonyms available to target applications. An example of the previously shown result of a REST based query to provide symptoms of urinary tract infections with associated read code bindings is shown in figure 5-17.
Figure 5-17 Evidence service reply describing a symptom collection for Urinary Tract Infection with associated read code bindings

5.8 Level 6: Learning, Versionable Electronic CPRs

The clinical evidence service at this point has been constructed based on manually created and curated evidence obtained from literature review. Updating evidence using a tool like Protégé on an on-going basis is not a sustainable approach in the long term. This addresses only the dissemination of evidence and ignores the derivation of evidence which is a crucial part of the Learning Health System. The clinical evidence that has been constructed must
also therefore support generation of evidence from electronic data sources to be considered a true LHS.

5.8.1 Data Mining and Evidence Population for UTI and Pyelonephritis

The potential for data mining electronic data sources to support research studies and production of clinical evidence has been demonstrated in a number of projects which utilised aggregated source of European primary care data to drive creation of research knowledge (32, 184).

I have collected data mined evidence that was derived using tools developed for the TRANSFoRm project. The TRANSFoRm approach developed a data mining application using an open source data mining tool called the Konstanz Information Miner (KNIME) that was used to produce quantified association rule combinations describing the relationships identified between ICPC2 coded diagnostic cues, demographic variables and diagnostic outcomes from the aggregated data sources (216, 244). This process provided empirically quantified diagnostic associations using calculated likelihood ratios (173, 245).

The TRANSFoRm project has used, with ethical approval granted as part of the project, anonymised primary care data captured from the Netherlands and Malta using the TransHIS EHR (177). This is a freely available open source EHR that was developed as part of the Transition project and captures International Classification of Primary Care 2 (ICPC2) coded values for RFE, diagnostic cues, demographic variables and outcome diagnoses. The TransHIS EHR structure links potentially many related clinical encounters in a single episode of care (EOC) structure that may occur over multiple time points as shown in figure 5-18.
KNIME has been used as part of the TRANSFoRm project to define workflows that pre-process the TransHIS record data and derive association rules based on ICPC2 codes. These rules identify all possible combinations of RfE, diagnostic cues and demographic variables (antecedent variables) that are linked with a recorded diagnostic outcome (consequent variable) as shown in the example in figure 5-19.

\[ \{\text{RfEs, Diagnostic Cues, Demographic Features}\} \rightarrow \text{Diagnosis} \]

\text{Antecedent Variables} \quad \text{--------} \quad \text{Consequent Variable}

\{\text{Abdominal Pain, Dysuria, Fever, Female}\} \rightarrow \{\text{Urinary Tract Infection}\}

ICPC2 Coded: \{D06, U01, A03, F\} \rightarrow U71

\text{Figure 5-19 Data mined association rule structure}

The patient records loaded into KNIME consisted of only the first patient encounter relating to each new episode of care for any patient. After cleaning (first encounter only from new episodes) 393,169 patient encounters were loaded into KNIME: 55,821 for Malta and 337,348 for the Netherlands. In total,
542,739 association rules were extracted from the data: 61,563 for Malta, 191,883 for the Netherlands, and 289,293 for both populations combined.

As part of the data mining process a number of quality measures were calculated that quantify the strength of association for the rule. For a full list of the quality measures and definitions refer to appendix E. These quality measures, such as the positive likelihood ration (LR+) and the negative likelihood ratio (LR-) with 95% confidence intervals provide a way of filtering rules to identify rules that have stronger levels of confidence than others.

The use of likelihood ratios also provides for the possibility of implementing Bayesian probabilistic reasoning to for diagnostic purposes as was described in section 2.6.2. According to Bayes Theorem, the post-test (posterior) odds of a specific diagnosis being made for a particular patient are equivalent to the pre-test odds multiplied by the likelihood ratios for the patient symptoms (LR) (246, 247). This involves the application of the calculated likelihood ratios for each diagnostic cue as multipliers to the pre-test probability for the condition (the prevalence in that specific population) to give a post-test probability for the condition that may have increased or decreased based on the presenting diagnostic cues for that individual patient. A full definition of the quality measures and data mining process is published by members of the TRANSFoRm project (172).

The distinct steps implemented in the data mining process are shown in figure 5-20 and are:

- derivation of association rules linking RfE, diagnostic cues and demographics to a recorded diagnoses made during the first encounter of a new episode of care
- calculation of association rule quality measures to determine the relative strength of each rule association derived
- filtering of association rules to allow selection of ‘high-quality’ association rules
- clinical review of selected rules to assess clinical validity of rules with respect to wider clinical body of evidence.
5.8.2 Clinical Association Rule Review

A web based association rule viewer allows curation and clinical review of all generated association rules from the KNIME tool (244). Selecting the ‘Rule Viewer’ menu item brings up the main rule display. The tool allows filtering of all rules generated from KNIME on any of the coded ICPC2 antecedent variables (RFEs or diagnostic cues/ (also called ‘anams’ here), and demographics)). The outcome diagnosis being examined can be filtered by selecting a consecutive variable.

In addition thresholds can be set on the defined quality measures to filter based on the strength of the rules required (support defined the number of cases for example). A scenario name can be entered that identifies a particular snapshot of rules run on a certain date. This allows multiple copies of rules at different points in time to be stored and retrieved based on a scenario label. The labels ‘NL_automated’ and ‘ML_automated’ can be used to retrieve data for Netherlands and Malta respectively as shown in figure 5-21.
The ‘filter’ button selects the required rules into the main rule viewer screen in the centre of the screen. By highlighting a particular rule the associated rule descriptions are shown at the bottom of the screen along with 95% confidence intervals. The column headings allow sorting of rules in ascending or descending order based on any of the quality measures as shown in figure 5-22.

Rules of interest can be selected for deployment to the ontology by selecting the ‘deploy’ button beside each rule on the main screen.
5.8.3 Rule Review and Curation in Practice

In this section we discuss an example of clinical review and single variable rule selection from the rule viewer tool. The tool was used to select single variable association rules of interest related to a consecutive diagnosis of U71 (urinary tract infection). A filter was applied to apply a cut off for the number of required patient cases for a rule using a ‘support’ value (the number of rule occurrences) > 5 cases. A filter was also applied to select only rules with a positive likelihood ratio greater than 2.

Using the 95% confidence intervals the rules were classified as follows based on a methodology previously used and published with TransHIS data (173):

- strong diagnostic predictors, in red were classified as LR+ > 8 with a tight confidence interval (value of observation is greater than width of interval)
- weak diagnostic predictors, in green were classified as LR+ >= 2 and LR+ <= 8 with a tight confidence interval (value of observation is greater than width of interval)
- diagnostic predictors for exclusion of diagnosis, in yellow were classified as LR+ <= 0.5.

Based on these classifications the strong predictors highlighted in red are shown in table 5-4 and were selected for deployment to the clinical evidence ontology. The table shows the LR values for Malta and Netherlands along with a combined pooled set of values. For comparison the LRs obtained from a JAMA systematic review of diagnostic predictors for Urinary Tract Infection are presented with these (248).
Table 5-4 Positive likelihood ratios for associated RFEs (label and ICPC code listed) and the episode title “UTI” in two populations.

<table>
<thead>
<tr>
<th>RFE Code</th>
<th>RFE Label</th>
<th>LR+ Netherlands</th>
<th>LR+ Malta</th>
<th>LR+ Combined</th>
<th>LR+ JAMA Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>U04</td>
<td>Incontinence urine</td>
<td>6.27 (5.04 – 7.80)</td>
<td>38.94 (15.93 – 95.2)</td>
<td>7.09 (5.74 – 8.76)</td>
<td></td>
</tr>
<tr>
<td>U07</td>
<td>Urine symptom/complaint, other</td>
<td>16.34 (11.51 – 23.21)</td>
<td>50.07 (21.99 – 114.01)</td>
<td>18.75 (13.59 – 25.87)</td>
<td></td>
</tr>
<tr>
<td>U05</td>
<td>Urination problems, other</td>
<td>6.35 (4.81 – 8.4)</td>
<td>6.85 (5.25 – 8.95)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L05</td>
<td>Flank/axilla symptom/complaint</td>
<td>2.08 (1.58 – 2.75)</td>
<td>2.23 (1.70 – 2.93)</td>
<td>1.1 (0.9 – 1.4)</td>
<td></td>
</tr>
<tr>
<td>U06</td>
<td>Haematuria</td>
<td>22.40 (18.63 – 26.94)</td>
<td>74.57 (49.54 – 112.26)</td>
<td>26.23 (22.19 – 31.01)</td>
<td>2.0 (1.3 – 2.9)</td>
</tr>
<tr>
<td>D06</td>
<td>Abdominal pain localized, other</td>
<td>2.59 (2.33 – 2.88)</td>
<td>7.67 (6.56 – 8.96)</td>
<td>3.09 (2.82 – 3.38)</td>
<td>1.1 (0.9 – 1.4)</td>
</tr>
<tr>
<td>U95</td>
<td>Urinary calculus</td>
<td>3.15 (1.38 – 7.22)</td>
<td></td>
<td>3.09 (1.35 – 7.05)</td>
<td></td>
</tr>
<tr>
<td>U29</td>
<td>Urinary symptom/complaint, other</td>
<td>17.16 (10.8 – 27.29)</td>
<td></td>
<td>18.17 (11.44 – 28.86)</td>
<td></td>
</tr>
<tr>
<td>U27</td>
<td>Fear of urinary disease, other</td>
<td>44.86 (38.7 – 52.02)</td>
<td></td>
<td>46.18 (39.89 – 53.45)</td>
<td></td>
</tr>
</tbody>
</table>
Table 5-4 (contd.) Positive likelihood ratios for associated RFEs (label and ICPC code listed) and the episode title "UTI" in two populations.

<table>
<thead>
<tr>
<th>RFE Code</th>
<th>RFE Label</th>
<th>LR+ Netherlands</th>
<th>LR+ Malta</th>
<th>LR+ Combined</th>
<th>LR+ JAMA Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>U01</td>
<td>Dysuria/painful urination</td>
<td>84.02 (77.87 – 90.67)</td>
<td>216.48 (180.68 – 259.39)</td>
<td>94.08 (87.70 – 100.93)</td>
<td>1.5 (1.2 – 2.0)</td>
</tr>
<tr>
<td>U71</td>
<td>Cystitis/urinary infection, other</td>
<td>185.80 (165.50 – 208.58)</td>
<td>305.98 (85.56 – 1094.27)</td>
<td>199.55 (177.81 – 223.95)</td>
<td>4.0 (2.9 – 5.5)</td>
</tr>
<tr>
<td>U02</td>
<td>Urinary frequency/urgency</td>
<td>44.65 (41.59 – 47.94)</td>
<td>151.66 (122.41 – 187.9)</td>
<td>50.78 (47.48 – 54.31)</td>
<td>1.8 (1.1 – 3.0)</td>
</tr>
<tr>
<td>A02</td>
<td>Chills</td>
<td>2.68 (1.37 – 5.25)</td>
<td></td>
<td>2.31 (1.35 – 3.96)</td>
<td></td>
</tr>
<tr>
<td>U13</td>
<td>Bladder symptom/complaint, other</td>
<td>36.49 (22.41 – 59.68)</td>
<td></td>
<td>39.24 (24.11 – 63.88)</td>
<td></td>
</tr>
<tr>
<td>U14</td>
<td>Kidney symptom/complaint</td>
<td>6.08 (2.59 – 14.3)</td>
<td></td>
<td>6.10 (2.61 – 14.30)</td>
<td></td>
</tr>
<tr>
<td>X15</td>
<td>Vaginal symptom/complaint, other</td>
<td>0.48 (0.26 – 0.9)</td>
<td>2.65 (1.18 – 5.95)</td>
<td>0.68 (0.42 – 1.12)</td>
<td></td>
</tr>
<tr>
<td>A03</td>
<td>Fever</td>
<td>0.81 (0.68 – 0.94)</td>
<td></td>
<td>0.72 (0.63 – 0.82)</td>
<td>1.6 (1.0 – 2.6)</td>
</tr>
<tr>
<td>X14</td>
<td>Vaginal Discharge</td>
<td>0.13 (0.06 – 0.32)</td>
<td>2.12 (0.86 – 5.13)</td>
<td>0.25 (0.14 – 0.47)</td>
<td></td>
</tr>
</tbody>
</table>
Table 5-4 (contd.) Positive likelihood ratios for associated RFEs (label and ICPC code listed) and the episode title “UTI” in two populations.

<table>
<thead>
<tr>
<th>RFE Code</th>
<th>RFE Label</th>
<th>LR+ Netherlands</th>
<th>LR+ Malta</th>
<th>LR+ Combined</th>
<th>LR+ JAMA Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>X15</td>
<td>Vaginal symptom/complaint, other</td>
<td>0.48 (0.26 – 0.9)</td>
<td>2.65 (1.18 – 5.95)</td>
<td>0.68 (0.42 – 1.12)</td>
<td></td>
</tr>
<tr>
<td>A03</td>
<td>Fever</td>
<td>0.81 (0.68 – 0.94)</td>
<td></td>
<td>0.72 (0.63 – 0.82)</td>
<td>1.6 (1.0 – 2.6)</td>
</tr>
<tr>
<td>X14</td>
<td>Vaginal Discharge</td>
<td>0.13 (0.06 – 0.32)</td>
<td>2.12 (0.86 – 5.13)</td>
<td>0.25 (0.14 – 0.47)</td>
<td></td>
</tr>
<tr>
<td>U01 and U02</td>
<td>Dysuria with Urinary Frequency</td>
<td>193.87 (165.54 – 227.05)</td>
<td>745.08 (431.150 – 1287.6)</td>
<td>222.67 (191.39 – 259.08)</td>
<td></td>
</tr>
</tbody>
</table>

LRs are highlighted according to the value (clinical significance) and reliability (95% CI). Strong predictors (LR+ >8 or LR- <0.2, CI width being equal to or smaller than the size of the observation itself) are in red. Weak predictors (LR+ >2-8, LR- 0.2-0.4, small CI) are in green. Associations with a wide CI (larger than the observation itself) or which are not clinically significant (LR+ <=2, LR- >=0.5) or have a CI which includes unity are not included.

The tool was then used to select single variable association rules of interest related to a consecutive diagnosis of U70 (pyelonephritis). The same criteria were applied to identify strong (red) and weak (green) predictors as summarised in table 5-5. The strong predictors highlighted that one of the key differentiators of UTI from pyelonephritis is the presence of flank pain.
Table 5-5 Positive likelihood ratios for associated RFEs (label and ICPC code listed) and the episode title “pyelonephritis” in two populations.

<table>
<thead>
<tr>
<th>RFE Code</th>
<th>RFE Label</th>
<th>LR+ Netherlands</th>
<th>LR+ Malta</th>
<th>LR+ Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>A02</td>
<td>Chills</td>
<td>57.97 (25.88 – 129.82)</td>
<td></td>
<td>33.49 (15.06 – 74.45)</td>
</tr>
<tr>
<td>U02</td>
<td>Urinary frequency/urgency</td>
<td>7.17 (4.31 – 11.94)</td>
<td></td>
<td>7.24 (4.34 – 12.05)</td>
</tr>
<tr>
<td>U70</td>
<td>Pyelonephritis/pyelitis</td>
<td>1001 (494.12 – 2027.87)</td>
<td>1001 (493.93 – 2028.62)</td>
<td></td>
</tr>
<tr>
<td>U06</td>
<td>Haematuria</td>
<td>17.38 (7.85 – 38.48)</td>
<td></td>
<td>16.57 (7.49 – 36.68)</td>
</tr>
<tr>
<td>U27</td>
<td>Fear of urinary disease, other</td>
<td>18.75 (9.84 – 35.75)</td>
<td></td>
<td>20.63 (10.82 – 39.34)</td>
</tr>
<tr>
<td>L05</td>
<td>Flank/axilla symptom/complaint</td>
<td>55.16 (41.65 – 73.06)</td>
<td></td>
<td>59.60 (44.97 – 78.99)</td>
</tr>
<tr>
<td>A03</td>
<td>Fever</td>
<td>11.81 (9.88 – 14.3)</td>
<td>4.19 (1.71 – 10.25)</td>
<td>9.00 (7.51 – 10.78)</td>
</tr>
<tr>
<td>U71</td>
<td>Cystitis/urinary infection, other</td>
<td>11.70 (7.15 – 19.13)</td>
<td></td>
<td>13.09 (8.00 – 21.42)</td>
</tr>
<tr>
<td>D01</td>
<td>Abdominal pain/cramps general</td>
<td>9.61 (5.98 – 15.46)</td>
<td></td>
<td>7.03 (4.38 – 11.31)</td>
</tr>
<tr>
<td>L02</td>
<td>Back symptom/complaint</td>
<td>7.41 (4.45 – 12.33)</td>
<td>17.44 (7.12 – 42.75)</td>
<td>7.00 (4.42 – 11.08)</td>
</tr>
</tbody>
</table>
Table 5-5 (contd.) Positive likelihood ratios for associated RFEs (label and ICPC code listed) and the episode title “pyelonephritis” in two populations.

<table>
<thead>
<tr>
<th>RFE Code</th>
<th>RFE Label</th>
<th>LR+ Netherlands</th>
<th>LR+ Malta</th>
<th>LR+ Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>D10</td>
<td>Vomiting</td>
<td>7.77(4.58 – 13.21)</td>
<td></td>
<td>4.95 (2.91 – 8.41)</td>
</tr>
<tr>
<td>D09</td>
<td>Nausea</td>
<td>6.79(4 – 11.54)</td>
<td>5.94(0.95 – 37.17)</td>
<td>5.75 (3.45 – 9.57)</td>
</tr>
<tr>
<td>L03</td>
<td>Low back symptom/complaint</td>
<td>2.33(1.34 – 4.04)</td>
<td></td>
<td>2.50 (1.44 – 4.34)</td>
</tr>
<tr>
<td>A06</td>
<td>Fainting/syncope</td>
<td>28.83(4.59 – 181.11)</td>
<td></td>
<td>0.98 (0.14 – 6.94)</td>
</tr>
</tbody>
</table>

As previously, LRs are highlighted according to the value (clinical significance) and reliability (95% CI). Strong predictors (LR+ >8 or LR- <0.2, CI width being equal to or smaller than the size of the observation itself) are in red. Weak predictors (LR+ >2-8, LR- 0.2-0.4, small CI) are in green. Associations with a wide CI (larger than the observation itself) or which are not clinically significant (LR+ <=2, LR- >=0.5) or have a CI which includes unity are not included.

The Dutch data and the combined dataset for ‘UTI’ (table 5-4) indicated that the presence of RFE’s ‘Cystitis/urinary tract infection’, ‘Dysuria’, ‘Fear of UTI’, ‘Urinary frequency/urgency’, ‘Haematuria’, ‘Urine symptom/complaint, other’ are all strong and reliable predictors for the diagnosis ‘Cystitis/Urinary Tract
Infection’. The RfE’s ‘Incontinence urine’, ‘Urination problems, other’,
‘Abdominal pain localised, other’, ‘Flank/axilla symptom/complaint’ are all less
strong, but reliable predictors for the diagnosis ‘Cystitis/Urinary Tract Infection’.
In the Dutch data the presence of RfE’s ‘Vaginal symptom/complaint’ or ‘Vaginal
discharge’ are strong but unreliable predictors to exclude a diagnosis of
‘Cystitis/Urinary Tract Infection’. The combined dataset indicated that ‘Vaginal
symptom/complaint’ was no longer a predictor for excluding a diagnosis of
‘Cystitis/Urinary Tract Infection’.

The Maltese data for ‘UTI’ (table 5-4) indicated that the presence of RfE’s
‘Dysuria’, ‘Urinary frequency/urgency’, ‘Haematuria’ are all strong, reliable,
predictors for the diagnosis ‘Cystitis/Urinary Tract Infection’. The RfE
‘Abdominal pain localised, other’ is a less strong but reliable predictor for the
diagnosis ‘Cystitis/Urinary Tract Infection’.

In table 5-5, the diagnostic associations for ‘pyelonephritis’ are analysed. The
Dutch data indicated that the presence of RfE’s ‘Flank/axilla
‘Abdominal pain/cramps general’ are all strong, reliable, predictors for the
‘Urinary frequency/urgency’, ‘Nausea’, ‘Abdominal pain localised, other’, ‘Low
back symptom/complaint’ are all less strong, but reliable predictors for the
diagnosis ‘Pyelonephritis’. The combined dataset resulted in a number of weak
predictors from the Dutch dataset becoming insignificant predictors. This loss of
significance is due to the smaller number of cases of pyelonephritis combined
from the Malta dataset. The Maltese data set did not present any clinically and
statistically significant predictors as only 83 cases of pyelonephritis were coded
which was not sufficient to generate the strength of evidence for the defined
quality cut-offs (number of cases > 5 and LR > 2). This reflects the smaller
Maltese population. By comparison there were 700+ cases identified in the
Netherlands population reflected in the stronger LRs produced and tighter
associated confidence intervals.
5.8.4 Association Rule Export

Once significant strong predictors in the form of association rules have been identified using the evidence review tool, these rules can then be selected for deployment to evidence service models from the main rule curation screen using the ‘rule sender’ menu. A list of selected rules is presented and selecting the ‘send’ option generates an XML output for the selected rules that can be saved onto the clinical evidence server and imported. An example of rules export generation is shown in figure 5-23. The XML schema for rules is shown in Appendix F.

<xml version="1.0" encoding="UTF-8">  
  - <Rule RuleScenario="NL_automated" RuleProvenance="Netherlands">  
    <RuleAntecedent Description="Dysuria/painful urination" AntecedentType="RFE">U01</RuleAntecedent>  
    <RuleConsecutive Description="Cystitis/urinary infection, other" ConsecutiveEncoding="ICPC2">U71</RuleConsecutive>  
    <RuleScore ScoreType="Support">1814.000</RuleScore>  
    <RuleScore ScoreType="Confidence">69.400</RuleScore>  
    <RuleScore ScoreType="Lift">28.910</RuleScore>  
    <RuleScore ScoreType="Specificity">0.997</RuleScore>  
    <RuleScore ScoreType="Sensitivity">0.234</RuleScore>  
    <RuleScore ScoreType="LR">-5.023</RuleScore>  
    <RuleScore ScoreType="LR">-6.768</RuleScore>  
    <RuleScore ScoreType="Odds">109.426</RuleScore>  
  </Rule>  
  - <Rule RuleScenario="NL_automated" RuleProvenance="Netherlands">  
    <RuleAntecedent Description="Flank/axilla symptom/complaint!" AntecedentType="RFE">U05</RuleAntecedent>  
    <RuleConsecutive Description="Cystitis/urinary infection, other" ConsecutiveEncoding="ICPC2">U71</RuleConsecutive>  
    <RuleScore ScoreType="Support">52.000</RuleScore>  
    <RuleScore ScoreType="Confidence">4.700</RuleScore>  
    <RuleScore ScoreType="Lift">2.020</RuleScore>  
    <RuleScore ScoreType="Specificity">0.997</RuleScore>  
    <RuleScore ScoreType="Sensitivity">0.007</RuleScore>  
    <RuleScore ScoreType="LR">-2.080</RuleScore>  
    <RuleScore ScoreType="LR">-2.080</RuleScore>  
    <RuleScore ScoreType="Odds">2.088</RuleScore>  
  </Rule>  
  - <Rule RuleScenario="NL_automated" RuleProvenance="Netherlands">  
    <RuleAntecedent Description="Urination problems, other" AntecedentType="RFE">U05</RuleAntecedent>  
    <RuleConsecutive Description="Cystitis/urinary infection, other" ConsecutiveEncoding="ICPC2">U71</RuleConsecutive>  
    <RuleScore ScoreType="Support">57.000</RuleScore>  
    <RuleScore ScoreType="Confidence">13.000</RuleScore>  
    <RuleScore ScoreType="Lift">5.657</RuleScore>  
    <RuleScore ScoreType="Specificity">0.999</RuleScore>  
    <RuleScore ScoreType="Sensitivity">0.007</RuleScore>  
    <RuleScore ScoreType="LR">-6.352</RuleScore>  
    <RuleScore ScoreType="LR">-6.352</RuleScore>  
    <RuleScore ScoreType="Odds">1.391</RuleScore>  
  </Rule>  
  - <Rule RuleScenario="NL_automated" RuleProvenance="Netherlands">  
    <RuleAntecedent Description="Urine frequency/urgency" AntecedentType="RFE">U02</RuleAntecedent>  
    <RuleConsecutive Description="Cystitis/urinary infection, other" ConsecutiveEncoding="ICPC2">U71</RuleConsecutive>  
    <RuleScore ScoreType="Support">1417.000</RuleScore>  
    <RuleScore ScoreType="Confidence">15.200</RuleScore>  
  </Rule>  
</RuleSet>
</xml>

Figure 5-23 – Generation of association rules export XML from the web rule sender
5.8.5 Ontology Rule Import and Update

The exported XML rule results of the data mining process can be imported into the general ontology model of evidence that was described previously and made accessible through the evidence clinical service. In order to do this an ontology updater tool has been developed to process the generated exported data mined rules.

The ontology updater tool is a command line tool developed as part of the TRANSFoRm project for the purpose of processing the generated XML rules and updating the clinical evidence ontology with evidence. XML files are saved into a specified directory on the evidence server and a command line tool is executed or can be scheduled to run periodically using the following command:

java –jar ontology_updater.jar [path to update file]

An illustration of the update process is shown in the figure 5-24.
First, the antecedents of each rule are examined to check if they describe a reason for encounter as indicated by the AntecedentType attribute in the XML document. In the event that an RFE is identified, the tool will check the clinical evidence ontology to ensure that it contains the item based on the associated
code binding. If not, then a new RFE is generated in the ontology and populated with the data from the antecedent. If the reason for encounter is found to exist, then the process simply continues scanning antecedents for RFEs until it reaches the end of the list for the given rule.

Similarly to RFEs in the antecedents, the consecutive is treated as an object in the clinical evidence ontology. Thus, if it cannot be found in the ontology based on its code binding, then the tool will generate and populate it using the data in the XML file. Once this is done, the tool is ready to update the quantification.

A quantification is defined by a unique combination of antecedents, consecutives and a country. If a quantification with the precise combination of these elements is not found in the ontology, then a new one is generated with a random identifier. This quantification is then linked with the relevant RFE objects and differential diagnosis objects in the ontology.

Upon completion of this task, the tool checks if there are any other rules to be processed. If so, the operation repeats. Otherwise, the tool will terminate its execution.

5.8.6 Rule Availability Through the Web Service

Rules imported into the ontology can then be accessed as ‘live’ evidence using the appropriate REST URL to retrieve the imported quantifications from the evidence service as shown in figure 5-25. This shows quantifications relating to dysuria as a diagnostic predictor of UTI as found in a Dutch population.
Comparisons with the Literature

A key objective of this analysis was to compare the consistency of the clinical associations generated from our analysis with previous high quality studies of clinical evidence relating to the two diagnostic conditions. As such high quality evidence based reviews or guidelines of clinical evidence supporting Urinary conditions were chosen for comparison (248-250).

A comparison of these predictors can be done against a gold standard literature review such as the JAMA review by Bent (248). This JAMA review identified 9 studies that describe quantified evidence in the form of likelihood ratios for using clinical symptoms, examination and history to diagnose urinary tract
infection in women. The methodological quality of the 9 studies was assessed and 5 of these were considered JAMA level 1 quality studies (see end of Appendix K for definitions of quality levels). In addition 6 of these studies were related to primary care or ambulatory settings and were therefore appropriate. On this basis this JAMA review was considered appropriate for comparison.

In the case of the JAMA review, the identified predictors compare favourably indicating similar strong predictors in the form of urinary frequency, haematuria and dysuria. Also agreeing with the JAMA review, self-labelling by patients was also shown as a strong predictor for UTI and the presence of vaginal discharge was considered a possible excluding factor.

A high level summary from the SIGN guidelines gives the following symptom based definitions of Cystitis and Pyelonephritis:

- **UTI** - 'evidence of urinary tract infection with symptoms suggestive of cystitis (dysuria or frequency without fever, chills or back pain)'
- **Pyelonephritis** - 'evidence of urinary tract infection with symptoms suggestive of pyelonephritis (loin pain, flank tenderness, fever, rigors or other manifestations of systemic inflammatory response)' (249).

The European Urology guidelines define Cystitis symptoms as 'Dysuria, frequency, urgency, pain or bladder tenderness'. These symptoms progress to Pyelonephritis with additional symptoms of 'Fever, Flank pain, Nausea, vomiting' (250).

The identified predictors from our analysis compare favourably with both the cystitis and pyelonephritis definitions. Our analysis indicated similar predictors in the form of urinary frequency, haematuria and dysuria from both population data sets for Cystitis. Other weaker predictors are consistent including abdominal pain or flank pain. Predictors for pyeonephritis such as fever, flank/back pain, nausea and vomiting were also consistent with literature.

In the Netherlands dataset, self-labelling by patients was also shown as a strong predictor for UTI and the presence of vaginal discharge was considered an excluding factor, also agreeing with findings from the JAMA review. Unlike the JAMA review, no association with fever was found for Cystitis and this is
consistent with later SIGN and European Urology guidelines which indicate this should be considered indicative of progression to Pyelonephritis. However we could not confirm any negative relationships between the presence of the symptoms ‘fever’, ‘chills’ or ‘back pain’ and a diagnosis of ‘UTI/Cystitis’. Our analysis also highlighted Cystitis itself as a significant predictor or Pyelonephritis indicating the relationship and progression of these conditions into each other.

The JAMA review with quantified likelihood ratios for specific cues concluded that ‘specific combinations of symptoms (e.g. dysuria and frequency without vaginal discharge raise the probability of UTI to more than 90%, effectively ruling in the diagnosis based on history alone’. In our analysis dysuria with frequency was found to be single biggest LR for a combination of cues and is consistent with the JAMA conclusions. Our calculated likelihood ratios were generally stronger than those from JAMA reviews which reflects the larger volumes of data analysed by us and this also reflects the effect of lower prior probability with the earlier presentation of illness in primary care as against emergency and secondary care.

5.8.8 Comparison between Populations

The number of associations and their relative strengths were found to improve with analysis of larger volumes of data as shown by the relative comparison of generated associations from Netherlands and Malta. The smaller volume of Malta data generated likelihood ratios that are larger and have wider confidence intervals. The prevalence of the condition has also shown to be important in requiring larger volumes of data as shown by the lack of predictors identified for the rarer Pyelonephritis found in Malta data. The key Cystitis indicators from Malta are consistent with the Netherlands data.

The results for the combined data (Malta and the Netherlands together) are heavily influenced by and agree with the Dutch data set as expected due to a larger number of patient encounters it contains. Where significant associations appear in the Dutch dataset without a comparable association in the smaller Mata dataset, this was reflected in some associations losing significance in the combined dataset (for example ‘vomiting’ in the case of Pyelonephritis).
5.8.9 eCPR Workbench Specification and Development

The availability of quantified evidence through the web service now allows construction of eCPRs based on that quantified evidence. An eCPR construction tool has been developed that allows the eCPR model to extract subsets of knowledge held in the general model of evidence to construct versioned, quantified eCPRs using the recognised formal eCPR structure described in level 3 of the eCPR maturity model. This adds normalised scoring schemes based on threshold approaches to decision making. These score schemes are defined based on the quantified evidence derived from the data mining process along with manual clinical review and interpretation of the general evidence and associated quality and measures.

The eCPR construction tool for constructing a data mined CPR for diagnosis of Urinary Tract Infection is shown in figure 5-26.

![Figure 5-26 eCPR construction tool to create electronic CPRs from quantified evidence obtained through a data mining process](image-url)
The tool allows for:

- browsing and selection of general data mined quantified clinical evidence that can be filtered based on condition, demographics or strength of evidence association
- definition of an eCPR by selecting eCPR elements based on a subset of the selected evidence
- definition of a user-friendly score for each eCPR element by rounding LR+ values to the nearest whole number
- definition of scoring risk levels along with clinical actions or interpretation of each score level.

The tool presents quantified data mined evidence (on the left hand side window) through access to the clinical evidence web service. This evidence can be filtered by RFE, diagnosis or demographics and elements can be selected as eCPR elements to construct versioned eCPRs using the recognised formal eCPR structure described in level 3. The middle window allows definition of the eCPR elements that form the rule and the right-hand side window allows definition of scoring scheme levels and associated clinical actions defined as text. This allows for definition of normalised scoring schemes based on threshold approaches to decision making. The score schemes risk levels and associated actions are defined through manual clinical review and interpretation of the general evidence and associated quality measures.

By selecting the data mined evidence that I previously presented for UTI and pyelonephritis, we can use the eCPR construction tool to construct the following data mined rules shown in table 5-6 and 5-7 and save them and make them available through the clinical evidence service. The data mined UTI rule compares favourably with existing CPRs for diagnosing UTI (248, 251).
**Table 5-6 Data mined eCPR supporting diagnosis of Urinary Tract Infection in primary care**

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dysuria/painful urination</td>
<td>5</td>
</tr>
<tr>
<td>Urinary frequency/urgency</td>
<td>2</td>
</tr>
<tr>
<td>Haematuria</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total Possible Score</strong></td>
<td><strong>8</strong></td>
</tr>
<tr>
<td>0-2 low risk / no antibiotic</td>
<td></td>
</tr>
<tr>
<td>3-4 medium risk / monitor over time</td>
<td></td>
</tr>
<tr>
<td>5-8 high risk / prescribe antibiotic</td>
<td></td>
</tr>
</tbody>
</table>

**Table 5-7 Data mined eCPR supporting diagnosis of Pyleonephritis in primary care**

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flank/axilla symptom/complaint</td>
<td>5</td>
</tr>
<tr>
<td>Fever</td>
<td>1</td>
</tr>
<tr>
<td>Cystitis/urinary infection, other</td>
<td>1</td>
</tr>
<tr>
<td>Abdominal pain/cramps general</td>
<td>1</td>
</tr>
<tr>
<td>Dysuria/painful urination</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>10</strong></td>
</tr>
<tr>
<td>0 – 3 low risk</td>
<td></td>
</tr>
<tr>
<td>4 – 5 medium risk</td>
<td></td>
</tr>
<tr>
<td>6-10 high risk</td>
<td></td>
</tr>
</tbody>
</table>

Saving the constructed eCPR makes it available through the web service and it can be accessed using a standard web based call from other applications:
http://<server>/ClinicalEvidenceRESTService/interfaces/query/cprs/DataMinedUTIRule

The XML output of the call is shown in figure 5-27.

```xml
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
  - <cprElement>
    - <cprElementCriteria>
      - <cprCriteriaScore>2</cprCriteriaScore>
      - <cprCriteriaPresent>true</cprCriteriaPresent>
    </cprElementCriteria>
    - <cprElementCue>
      - <code>IA1..12</code>
        - <codingScheme>ReadCodeBinding</codingScheme>
        - <description>Polyuria</description>
        - <isPrimary>false</isPrimary>
    </cprElementCue>
    - <code>Blindings>
      - <code>IA1..13</code>
        - <codingScheme>ReadCodeBinding</codingScheme>
        - <description>Urinary frequency - high</description>
        - <isPrimary>true</isPrimary>
    </code>Blindings>
    - <cprElementNumber>1</cprElementNumber>
  </cprElement>
  + <cprElement>
  + <cprElement>
    - <cprScoreScheme>
      - <CPRSscoreDecision>No antibiotic</CPRSscoreDecision>
      - <CPRSscoreLevelEnd>2</CPRSscoreLevelEnd>
      - <CPRSscoreLevelStart>0</CPRSscoreLevelStart>
      - <CPRSscoreRisk>Low</CPRSscoreRisk>
    </cprScoreScheme>
    - <cprScoreScheme>
      - <CPRSscoreDecision>Monitor over time</CPRSscoreDecision>
      - <CPRSscoreLevelEnd>4</CPRSscoreLevelEnd>
      - <CPRSscoreLevelStart>3</CPRSscoreLevelStart>
      - <CPRSscoreRisk>Medium</CPRSscoreRisk>
    </cprScoreScheme>
    - <cprScoreScheme>
      - <CPRSscoreDecision>Prescribe antibiotic</CPRSscoreDecision>
      - <CPRSscoreLevelEnd>7</CPRSscoreLevelEnd>
      - <CPRSscoreLevelStart>5</CPRSscoreLevelStart>
      - <CPRSscoreRisk>High</CPRSscoreRisk>
    </cprScoreScheme>
    <cprName>DataMinedUTIRule</cprName>
  </cprElement>
</clinicalPredictionRule>
```

Figure 5-27 A Web service call to a data mined eCPR (example of Level 6 learning, versionable eCPR)

This demonstrates the combined use of both the developed ontology models and clinical evidence service along with the generation of quantified data mined evidence that can be used to construct eCPRs. This provides a platform for
third party applications to access data mined evidence in the form of a general
model of evidence or more specifically in the form of eCPRs. Up to this point
however we have shown individual calls to discreet pieces of diagnostic
information. What is now needed is a demonstration of the use of the service as
part of diagnostic workflow. I will now discuss how the technical platform can be
used to implement diagnostic strategies along with an example of practical use
from a third party diagnostic decision support application developed separately
for the TRANSFoRm project.

5.9 Workflow Integration Formalising the Diagnostic Process
In the literature review I highlighted some diagnostic strategies and their
associate heuristic limitations that can inform development of a technical
solution to diagnosis. I also stressed the importance of placing any such
technical solution in the context of a defined diagnostic workflow.

The necessary workflow can be described as the interactions between the EHR,
the evidence service and the decision support tool itself. The initial trigger for
the process is the entering of a patient RFE which results in an initial list of
differential diagnoses to consider, filtered based on the initial patient details
extracted from the EHR. The TRANSFoRm project identified a two-phase
approach to describe the workflow for diagnostic decision support. The phases
are referred to as phase 1 ‘suggesting’ and phase 2 ‘alerting’ and involve
interaction between the clinician, the clinician EHR system (with integrated
decision support tool) and the clinical evidence service as shown in figure 5-28
(252).
Figure 5-28 TRANSFoRm 2 phase diagnostic process

The first ‘suggesting’ phase is triggered by the presenting complaint(s), or RFE. A number of differential diagnoses may be associated with any presenting reason for encounter. These can then be filtered based on core demographics that are obtained from the electronic health record (age, sex, ethnicity, nationality) and underlying patient risk factors available from the electronic health record (lifestyle or clinical history). The output from this phase is a suggested list of differentials to consider based on the underlying patient reason for encounter, along with all supporting diagnostic cues to consider for each differential, before any detailed patient consultation has taken place.

The second ‘alerting’ phase will involve more detailed diagnostic workup by allowing the clinician to gather and record diagnostic cues as part of the patient consultation. As will be shown later in the implementation in section 5.10.2, the evidence is presented for all of the differentials under consideration for the RFE in a prompted fashion allowing the clinician to drill into any differential to see supporting evidence to consider. This facilitates a bottom-up consideration of the evidence based on diagnostic cues and working up to a differential to consider. Alternatively the evidence can be considered in a top-down fashion by looking at the presenting evidence for differential diagnosis and working down to supporting diagnostic cues, This is done with a view to limiting confirmation
bias by allowing the clinician to see all the supporting evidence to consider early in the consultation rather than allowing them to pursue an initial one or two diagnoses in isolation and searching for diagnostic cues to support that clinical hypothesis.

An interactive dialogue takes place between the decision support tool and the underlying clinical evidence service. The gathered patient consultation diagnostic cues are compared against the clinical evidence available for each diagnosis. This patient evidence set is submitted to the evidence service resulting in a filtering process of differentials to consider that is returned back to the decision support interface. The key output from this process will be a ranked list of competing diagnoses ordered by relative likelihood in the form of a Bayesian post-test probability along with consideration of the disease severity to highlight ‘not to miss’ diagnoses. After a working diagnosis is selected, all of the gathered consultation diagnostic cues are recorded back to the EHR to support a feedback loop for future data mining.

Both ‘suggesting’ and ‘alerting’ strategies involve interaction between the evidence service, the decision support tool interface and the local EHR. The approach described here provides for the delivery of the clinical evidence in a highly structured and open standards based format with support for local coding schemes. This workflow process and the interactions between the EHR, the decision support tool interface and the evidence service as implemented in TRANSFoRm are summarised in figure 5-29.

It should be noted that implementation of this workflow does require integration work to be done to map the XML generated evidence service output format to the required programming interfaces used by the local EHR but it is significantly easier through use of open standards used in the clinical evidence service. This could also be done through the use of an integration engine or a common data integration model (CDIM) to which local EHR formats can be mapped as was developed in TRANSFoRm (226). Mapping of evidence therefore needs to consider both mapping of concepts and mapping of terminology coding schemes (226, 253).
The ontology of evidence developed as part of this research was deliberately developed at a high level of conceptual granularity that results in a small number of core clinical evidence concepts (RFE, diagnosis, cues, demographics etc.). As such the effort to map from these evidence concepts to local EHR data fields that act as a trigger sent to a local EHR interface can be measured in days rather than weeks.

A more time consuming consideration is the use of terminology codes in different local EHR systems that need to be attached to the clinical evidence model. In this research the evidence code bindings that were populated into the clinical evidence models targeted both ICPC2 and NHS read codes. These can therefore be used against any target local EHR that uses those either of those coding schemes. For other EHRs that use different coding schemes the additional terminology code bindings to support it needed to be defined.

If a data mining source is used that has the appropriate terminology codes then the evidence ontology can be automatically populated to bring in the appropriate code bindings (as was demonstrated here using TRANSHIS data that was based on ICPC2 codes). Alternatively the code bindings would have to be manually entered into the clinical evidence ontology, a time consuming task that based on the experiences of this research can be measured in weeks of effort (as was the case for NHS read codes in this research work). In this situation an alternative would be to make use of a dedicated terminology server which has mappings between different terminology sets (226, 253). This could be used to do a lookup and map to other appropriate coding schemes either at runtime or as a batch process to enrich the evidence service with additional terminology content. This work has focussed on the backend evidence representation and EHR integration was limited to a single local system. A terminology server was therefore not implemented. Such an approach would provide a more flexible approach to evidence integration to other EHR systems.
Figure 5-29 TRANSFoRm diagnostic workflow and interaction of decision support components
5.9.1 ‘Suggesting’ Phase Diagnostic Strategies

The following diagnostic strategies may also be identified as diagnostic strategies used in the suggesting phase. In addition to a brief description, the key ontology concepts and features that can be used to implement the strategy are highlighted (254).

Presenting Complaint - the patient expresses a general expression of the underlying reason for their clinical presentation e.g. ‘doctor I have a pain in my abdomen’. The reason for encounter (RFE) is captured as a formal concept in the general model of clinical evidence. The RFE may be considered a general presenting problem or a defined diagnosis label. This RFE concept is associated through ontological relationships (hasDifferentialDiagnosis) with defined differential diagnoses to consider.

Self Labelling – the patient expresses a definition of the problem that may also be a diagnostic label e.g. ‘doctor I think I have a urinary tract infection’. The reason for encounter (RFE) is captured as a formal concept in the general model of clinical evidence. Where the patient expresses the RFE as an actual diagnostic label (rather than a presenting problem) this may indicate that the patient has already suspected and self-labelled the condition that they may be suffering from. In the case of urinary tract infection for example I previously showed using data mined evidence that self-labelling has high quantifiable value for the condition.

Pattern Recognition Trigger – identification of a potential differential diagnosis to consider based on diagnostic pattern recognition and experience on the part of the clinician. This will involve the identification of a potential diagnosis using a collection of diagnostic cues that support the presence or absence of a condition. The ‘Diagnosis’ ontology concept records associated diagnostic cues (‘Cue’) in the form of ‘Sign’, ‘Symptom’, ‘Risk’, ‘Test’. Explicit ontological relationships link collections of cues e.g. ‘UrinaryTractInfection’ ‘hasSymptom’ ‘Dysuria’.

‘Spot’ diagnosis – potential diagnosis is identified primarily based on patterns of visual or auditory patterns of cues – e.g. Chest Crackles. The ‘Sign’ concept
allows for capture of clinician observed signs that could be auditory or visual in nature e.g. Weight Loss.

5.9.2 ‘Alerting’ Phase Diagnostic Strategies (254)

Restricted Rule Outs (Exclusion) – exclusion is a useful diagnostic strategy to rule out potential diagnoses that may not fit a patient case. The utility of supporting negative ontological statements as well as positive associations in our ontology allows for this e.g. ‘IsNotSymptomOf’, ‘IsNotSignOf’, ‘IsNotRiskOf’, ‘IsNotTestOf’. For example we saw previously in the data mined evidence that vaginal bleeding can be grounds for a restricted rule out of urinary tract infection.

Probabilistic Reasoning – probabilistic reasoning may be implemented as a technological diagnostic strategy as previously discussed. The ‘Quantification’ and ‘Context’ concepts allow for the capture of measures of the strength association of diagnostic cues to related diagnostic conditions (within the context of a specific associated population demographic). This may be used to implement probabilistic reasoning represent likelihood ratios for individual cues

Pattern Recognition Fit - identification of a potential differential diagnosis to consider based on diagnostic pattern recognition and experience on the part of the clinician. This will involve the identification of a potential diagnosis using a collection of diagnostic cues that support the presence or absence of a condition. The ‘Diagnosis’ ontology concept records associated diagnostic cues (‘Cue’) in the form of ‘Sign’, ‘Symptom’, ‘Risk’, ‘Test’. Explicit ontological relationships link collections of cues e.g. “UrinaryTractInfection” “hasSymptom” “Dysuria”.

Clinical Prediction Rule – the role of CPRs as a diagnostic strategy was discussed previously. CPRs are explicitly supported by a separate CPR ontology model. Concepts include ClinicalPredictionRule, CPRElement and ClinicalPredictionRuleScore to allow for the representation of computable CPRs based on the diagnostic cues obtained from the general model of evidence. CPRs are related to specific diagnoses through the ‘hasCPR’ ontology relationship.
Red Flag Cues – specific combinations of strongly associated diagnostic cues may give rise to a stronger belief in the presence of a particular diagnosis (255). An example are the NICE clinical guidelines that identify 4 key red flags for the detection of ovarian cancer in primary care (256). The ontology has also implemented a ‘RedFlagGroup’ concept to associate ‘Cues’ in groups of red flags for a particular condition. The relationships ‘hasRedFlagCues’ captures the relationship of red flag grouping with a particular diagnosis e.g. ‘OvarianCancer’ ‘hasRedFlagCues’ ‘OvarianCancerRedFlags’. The red flag grouping is associated to the individual cues themselves through ‘hasSign’, ‘hasSymptom’ etc.

Severity of conditions – certain diagnoses are inherently more severe or life-threatening to the potential safety of a patient. It is therefore useful to record an ontological relationship of ‘hasDiseaseSeverity’ against each diagnosis in the ontology. Ovarian cancer would have an associated disease severity of ‘High’. These disease severities can be used to classify ‘not to miss’ diagnoses when presenting differential diagnoses for consideration.

5.10 Evidence Service Evaluation in Practice
An evaluation of this work will be done based on:

- demonstration of an actual implementation of an independent third party decision support tool development that utilised the evidence service and evaluated its performance with a sample of GPs in the UK as part of the TRANSFoRm project
- a comparison to technical best practice in diagnostic decision support development as described in a reviews of desirable architectural features for ensuring successful decision support (93, 128).

5.10.1 Development of Diagnostic Tools Using the Clinical Evidence Service
The TRANSFoRm project developed a fully functional diagnostic decision support tool that was integrated with an EHR for use in family practice. This decision support tool was developed using the C# language and a client layer REST library was developed to allow the tool to access the clinical evidence
service as described in this work as was previously shown in figure 5-9. The REST client layer in effect implements and coordinates the dialogue between the tool and the clinical evidence service through a series of clinical evidence calls that implement the diagnostic strategies and workflow that was described in figure 5-28.

The calls that are sent to the clinical evidence service are based on the content of patient data that is submitted by the tool to the evidence service. The REST client handles exchange of patient data between the decision support tool with appropriate calls sent to the backend evidence service. The client accepts patient data in the form of a XML patient evidence set describing the patient RFE, demographics (extracted from the EHR) and the underlying cues confirmed through consultation with the patient (extracted from the integrated decision support tool itself) shown in figure 5-30.

The decision support tool is implemented using C# and hence a decision was made to implement the client layer in C# also for ease of integration. In order to
implement the diagnostic workflow previously described, an evidence client
dynamic linked library (DLL) has been created and can be linked into the
decision support tool interface. This DLL implements a number of classes and
methods that provide C# wrappers for the REST based web service calls
previously described. It should be pointed out that the C# client library can just
as easily be ported to another language such as Java as tools exist to generate
the required REST calls automatically in Java development tools. For a full list
of the methods and C# classes implemented refer to the following on-line
documentation:

http://phaedrus.scss.tcd.ie/munnellg/

The most important class used by the decision support tool is the
ClinicalEvidenceWebInterface class which provides the methods that can be
called by the decision support interface itself. These classes and methods are
summarised in Appendix G and H. In addition a PatientCase class accepts an
XML serialisable C# PatientEvidenceSet that allows the CDSS interface to
submit or update the exchange of a set of patient evidence obtained from a
combination of the EHR and the decision support interface, using the
‘UpdatePatientCase’ method. This method returns a ranked list of differentials
to consider listed as ‘most likely’, ‘not to miss’ and ‘others’. The PatientCase
class and a portion of the C# PatientEvidenceSet class are shown Appendix I
and J.

5.10.2 Evidence Service Integration with TRANSFoRm Diagnostic Decision
Support Tool

The evidence service provides the decision support tool with ontology driven
coded prompting and recording of coded patient diagnostic cues as shown in
figure 5-31. The diagnostic decision support tool is embedded and interoperable
with an EHR in family practice using the EHR programmable interface (the In
Practice Systems Vision 3 EHR). The clinical motivation and design process for
this tool has been described fully elsewhere (252, 257).
The tool allows bottom-up input of observed or absent patient cues independent of associated diagnosis (left window) or top-down drilling into and selection of evidence cues supporting specific diagnoses (right window). The evidence service returns recommendations in the form of a dynamically updated ranked list of differentials to consider by keeping a cue count for each differential under consideration.

This list is based on the presenting RFE and ordered in descending cue count based on the total number of patient cues confirmed present for each differential along with the supporting underlying evidence cues for each diagnosis. As each patient cue is selected the cue counts are recalculated. The differentials are re-ranked in descending order based on the cue count. In addition each diagnosis has a prevalence category assigned to it (common, uncommon and rare). An interactive and iterative diagnostic conversation can take place between the third party consumer as presence or absence of patient cues are confirmed, appropriate patient contextualised REST queries are executed and the re-ranked diagnosis list is supplied to the consumer tool.
Where differentials have the same number of cues present they are ordered by highest prevalence first. These are used to dynamically rank potential differential diagnoses to consider (most likely at top) based on the patient presenting RFE along with the evidence supporting each diagnosis under consideration. Upon exiting the tool a working diagnosis can be confirmed and the coded evidence cues and current working diagnosis can be saved back and recorded for future reference in the patient EHR. The decision support tool is therefore integrated with the EHR itself allowing a full set of consultation cues and decision support recommendations (in the form of a confirmed diagnosis) to be recorded and made accessible from the EHR tool itself.

The successful development of the TRANSFoRm tool demonstrates that the clinical evidence service architecture:\(^{(229)}\)

- works in practice and was clinically shown in a separate evaluation to be effective in helping to improve diagnostic decision making in practice as reported in the TRANSFoRm deliverable for the validation where 'mean diagnostic accuracy was higher with the CDSS than without: means 0.58 [95% CI 0.52-0.65] vs. 0.50 [95% CI 0.42-0.57]; odds ratio 1.41 [95% CI 1.13-1.77] (P<0.01)' \(^{(258)}\)
- demonstrates the necessary technical flexibility to be incorporated into the clinical workflow of family practitioners \(^{(30, 229)}\).

This demonstration of the practical implementation of the clinical evidence service allows us to conclude that the clinical evidence service works and has practical applications in a particular implementation. This is a minimum requirement for successfully evaluating the architecture. A more general consideration needs to be given to comparing the architecture against current best practice in this area. I will now consider more generally an evaluation of the architecture based on reviews of desirable features highlighted for successful decision support architectures as found in the literature.
5.10.3 Evaluation Based on Review of Desirable Features for Decision Support Systems

Wright et al. carried out a review of decision support systems (SEBASTIAN, SAGE) and decision support languages (GLIF, GELLO, Arden) and developed a framework for evaluating decision support architectures. This allows comparison with the study architecture of desirable functions of decision support initiatives that have been deemed leaders in this area, shown in figure 5-32 (128).

Table 5-8 Desirable functionality for implementing clinical decision support systems amended from Wright et al.(128)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Stand Alone</th>
<th>Integrated</th>
<th>Arden</th>
<th>GLIF 3.5 /GELLO</th>
<th>SEBASTIAN</th>
<th>SAGE</th>
<th>This Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avoids vocabulary issues</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shareable</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can view decision support content</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content separate from code</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automatic central updates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content integrated into workflows</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supports event driven CDS</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supports non-event driven CDS</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supports decision support over populations</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enables separation of responsibility</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enables composition of rules</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allows black-box services</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fee choice of knowledge representation syntax</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Going through this list of features I will indicate if the functionality is ‘satisfied’, ‘partially satisfied’ or ‘absent’ as demonstrated in the course of this research work.

**Avoids vocabulary issues - (satisfied)** – the decision support system should avoid conflicts with terminology and vocabulary. The clinical evidence service specifically addresses this through the representation of clinical evidence independently as an ontology with explicit support for binding of vocabulary terms to that evidence. Any desired vocabulary or terminology may be used for binding as was discussed in section 5.7.

**Shareable – (partially satisfied)** – the knowledge base should be shareable across many systems. This is explicitly satisfied through provision of a separate clinical evidence service and support for openly available technical standards as demonstrated in the TRANSFoRm project and was discussed in section 5.6. As previously noted though some local mapping to local EHR formats may still be required to integrate fully with a local EHR.

**Can view decision support content – (satisfied)** - decision support knowledge can be viewed in a number of ways using this architecture. Queries can be run directly using the Sesame triple store query language (SPARQL) discussed in section 5.5.3. Queries can also be run using the web service itself and obtained in XML format as demonstrated in section 5.6. To generate human interpretable versions of the XML it will be relatively easy (but not done here) to generate an application that applies XML formatting using stylesheets to present the content as a document. Data mined rule context is available through the web based data mining tool that was previously described in section 5.8. Finally manually curated knowledge could be viewed using the source ontology file through the Protégé modelling tool described in section 5.5.

**Content separate from code – (satisfied)** – The explicit model based approach using an ontology as described in section 5.5 ensures that clinical evidence content is separate from any coding done and implemented in the client layer.
**Automatic central updates** – *(satisfied)* – The discussion of the data mining approach with support for data mined rule export and import into the centralised clinical evidence service (section 5.8.3) demonstrates support for centralised updates of evidence. Once updates are made, all client decision support tools that consume the service can receive the updates.

**Content integrated into workflows** – *(partially satisfied)* – The ability to programmatically combine evidence service calls as part of a client layer that implements a diagnostic workflow was demonstrated in sections 5.9 and 5.10.3. The decision support tool was fully integrated with an EHR and the recommendations and consultation cues are saved to and made available in the EHR itself as described in section 5.10.2. The decision support service is integrated with the decision support tool, and the decision support tool is integrated with the EHR. An open technical interface was used to integrate the decision support service and decision support tool. However a non-standard interface was used to integrate the decision support tool and the EHR. On that basis we can say that workflow integration was partially successful but demonstrates integration across all required decision support components. More work is needed to standardise the integration of the EHR using open technical standards.

**Supports event driven clinical decision support** – *(satisfied)* – The discussion of the TRANSFoRm decision support tool in 5.10.3 demonstrated an event driven form of decision support based on entering a presenting reason for encounter (RFE) that triggers calls to the backend evidence service.

**Supports non-event driven clinical decision support** – *(satisfied)* – The ability to trigger a decision support process based on a batch scheduled process has not been explicitly demonstrated as part of this work. The architecture however can support such a process through the Jersey REST based interface and the Sesame API that allow programmatic update or querying of the knowledgebase (as shown in Appendix B or Appendix C). In the example of TRANSFoRm the client layer implemented a library that was incorporated into an interactive decisions support interface. Alternatively the client layer could be a command line driven program in its own right that can be
executed as a batch process and scheduled to periodically run over a population of patients from an EHR repository.

**Supports decision support over populations – (satisfied)** – As per the discussion on non-event driven clinical decision support the architecture can support batch processing over a population of patients (rather than a single patient triggering a single event).

**Enables separation of responsibilities – (satisfied)** – This functionality specifically addresses the need to separate programmer responsibilities from clinical content curation responsibilities. These requirements have been explicitly provided for through the separation of clinical content in the form of models (in section 5.5) and the provision of curation tool for reviewing evidence before import (in section 5.8)

**Enables composition of rules – (satisfied)** – This functionality is specifically related to the ability of one decision support system being able to call the rules and content available from another decision support system. The service oriented approach and support for open technical standards demonstrates the feasibility of this as described in section 5.10.

**Allows black-box services – (absent)** – This functionality allows for the hiding of rule content to consumers on the basis that the evidence content may be proprietary in nature. The approach I have discussed explicitly takes an ‘open’ approach to making evidence content available so all content can be seen through simple queries to the REST based service. It is arguable that the ability to ‘hide’ decision support rule content from consumers is not necessarily a desirable feature in its own right as it may hinder trust in the underlying content.

**Free choice of knowledge representation syntax – (satisfied)** – This functionality addresses whether users of the decision support functionality need to have an explicit knowledge of the underlying proprietary knowledge representation format in order to use it. In the case of this architecture the underlying knowledge is represented as an ontology but is made available through the web service which can return results in XML, JSON and RDF
formats (as described in section 5.6). These formats are all based on open and widely used technical standards that are not proprietary in nature.

Kawamoto et al. provided another review identifying four statistically significant features for ensuring clinically effective decision support. They reviewed the literature describing decision support systems and split them between those that demonstrated statistically significant improvements in clinical practice with those reporting no improvement (93). Support for decision support features was compared between the two groups. Features were identified using univariate analysis that were statistically associated with successful implementations and absent from unsuccessful implementations as shown in figure 5-33.

<table>
<thead>
<tr>
<th>Feature*</th>
<th>Adjusted odds ratio (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary analysis (all CDSS, n=71)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automatic provision of decision support as part of clinician workflow</td>
<td>112.1 (12.9 to ∞)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Provision of decision support at time and location of decision making</td>
<td>15.4 (1.3 to 300.6)</td>
<td>0.0263</td>
</tr>
<tr>
<td>Provision of recommendation rather than just an assessment</td>
<td>7.1 (1.3 to 45.6)</td>
<td>0.0187</td>
</tr>
<tr>
<td>Computer based generation of decision support</td>
<td>6.3 (1.2 to 45.0)</td>
<td>0.0294</td>
</tr>
</tbody>
</table>

*Figure 5-32 Figure 5 30 Desirable functionality for implementing clinical decision support systems courtesy of Kawamoto et al.(93)*

**Automatic provision of decision support as part of clinician workflow** – *(partially satisfied)* – As was discussed for Wright et al, this requirement is partially satisfied.

**Provision of decision support at time and location of decision making** – *(satisfied)* – The TRANSFoRm decision support tool demonstrated diagnostic decision support as part of an EHR used during the family practice consultation. The use of the tool significantly did not impact on the total time for consultation (258)
Provision of recommendation rather than just an assessment – (satisfied) – Whilst the focus of this work is exclusively on diagnostic decision support, the support for representation of CPRs allows for provision of clinical recommendations based on the threshold approach to risk assessment. The example of the Alvarado score that was discussed in section 5.3.1 specifically includes clinical recommendations in response to the risk assessment of appendicitis.

Computer based generation of decision support – (satisfied) – The architecture is inherently a computerised and computable form of decision support.

5.10.4 Interpretation of Architecture Evaluation

In conclusion it can be stated that an evaluation of the architecture developed in this research scored highly in support for desirable features compared to other existing approaches. All features were demonstrably present with the exception of being able to hide clinical content. As stated it is arguable as to whether this is a desirable feature to develop trust in decision support systems.

A practical demonstration of the open technical nature and feasibility of the architecture being used in practice has also been provided through the development of the TRANSFoRm decision support tool (30, 229). The TRANSFoRm decision support tool demonstrated applying evidence based on the general model of evidence described in section 5.2.10. The eCPR model of evidence was not deployed using the tool but this can be accessed for use by third party tools in exactly the same way using the same clinical evidence web service and technical infrastructure as was done for accessing the general model of evidence. As was previously discussed in section 4.3.4, the eCPR model is a more structured view of the clinical evidence content contained in the general model of evidence. On this basis it is stated that the architecture succeeds in adhering to current best practices relating to decision support design and extends it by placing its use in the context of a learning health system and demonstrating support for data mining evidence and eCPRs as described in section 5.8.
6 Discussion / Conclusion

The traditional focus on literature based derivation and narrow validation of CPRs has severely limited their wider acceptance. This research has given appropriate consideration of technology-based strategies that encourage wider scale derivation and dissemination of eCPRs. This research addresses the need to put in place a suitable delivery mechanism that allows dissemination of eCPRs as part of clinician workflow at point of care using the EHR. Addressing the difficulties in progressing CPRs to validation and impact analysis was not the primary focus of this work but has been acknowledged as important in the literature review. An ICT based approach will not enable eCPR validation or impact analysis on its own but provides an alternative more dynamic eCPR lifecycle and intervention delivery mechanism with which it might be facilitated through dissemination to potentially larger patient populations.

The LHS appears to be an obvious technical context within which this might be implemented and the research questions of this work were posed with a view to practically testing this in detail. In this section I will discuss what conclusions can be drawn about the use of computable eCPR models as part of an LHS with a view to improving the CPR development process. Specifically I will consider:

- the degree to which the original research questions have been answered
- the practical limitations of this work that were observed as part of its actual implementation
- future work and identification of areas of research that might be done to address the limitations described.

6.1 Delivery of Research Question Implementation Aims

At the outset of this work in section 1.1. and summarised again here for ease of reference, I outlined three research questions to be addressed relating to a number of identifiable gaps in the literature. The research questions were:
• Research Question 1 - Can traditional literature based CPRs be represented in a computable, generalisable and interoperable format using open technical standards?

• Research Question 2 - Can clinically valid eCPR knowledge be dynamically developed using aggregated sources of primary care electronic health record data?

• Research Question 3 - Can a Learning Health System infrastructure support the derivation and dissemination of eCPRs as a service for use by third party tools providing diagnostic decision support?

Each of these questions will be discussed in the context of establishing how successful the research has been in answering them and delivering the implementation and conceptual aims identified for each one.

In relation to research question 1 it has been demonstrably shown that CPRs can be represented electronically using flexible open and technical standards described in sections 5.4 and 5.5. In addition a general structure and model representation has been described that captures the general form for representation of any diagnostic CPR. Importantly the technical approach to delivering this is portable and has been shown to adhere to best practice with respect to decision support tool design as evaluated in section 5.10.3. Semantic interoperability has been supported through flexible support for terminology coding and system interoperability (sections 5.5 and 5.7) and has been demonstrated with successful integration with a third party decision support tool consumer of the service (section 5.10.3). The implementation aims for this research are shown in table 6-1 and were successfully delivered:
Table 6-1 Implementation aims of research question 1

<table>
<thead>
<tr>
<th>Implementation aim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop a generic model for the representation of computable eCPRs</td>
</tr>
<tr>
<td>Develop an infrastructure based on openly available tools and technology standards based on proven implementations used in other commercial domains</td>
</tr>
</tbody>
</table>

With respect to research question 2 I conclude that derivation of clinically meaningful eCPRs from aggregated sources of coded patient data is feasible albeit in the context of a limited clinical scenario. A defined data mining and evidence curation process has been demonstrated in section 5.8, within which data mined diagnostic eCPRs were successfully constructed and disseminated using developed electronic tools. Importantly these data mined rules were shown to be consistent and agree well with existing clinical literature and CPR guidelines derived using the traditional CPR development lifecycle. The broader utility of such an approach is however heavily dependent on the wider availability of good quality EHR data with suitably expressive terminology coding. A more flexible and interoperable method of connecting EHR data to the data mining modules is needed. It can be stated therefore that the success and wider generalisability of this approach is subject to a number of caveats. These caveats reflect limitations of the data mining approach and will be discussed in more detail later. The implementation aims shown in table 6-2 for research question 2 were met in the context of one electronic repository applied to a limited clinical scenario:

Table 6-2 Implementation aims of research question 2

<table>
<thead>
<tr>
<th>Implementation aim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploit existing repositories of electronic patient data for the purposes of demonstrating how clinically valid eCPRs can be electronically derived in practice</td>
</tr>
<tr>
<td>Validate the electronically derived eCPR against existing clinical literature to assess its agreement and consistency with current clinical knowledge</td>
</tr>
</tbody>
</table>
Successfully answering research questions 1 and 2 along with the development of a conceptual framework for implementing eCPRs defined in section 4, allow us to conclude that research question 3 is satisfied by demonstrating dissemination and the derivation of eCPRs as part of a single LHS like implementation. Based on this work it can be concluded that the derivation, curation and dissemination of eCPR knowledge is demonstrably achievable using a distributed technical platform developed on open technical components.

This platform has demonstrated open integration with third party decision support consumers and could be extended beyond integration with the EHR in forms such as mobile devices for example. I conclude that the LHS is a very good conceptual fit for describing and supporting the electronic derivation and dissemination of eCPRs. The fact that eCPRs can be disseminated electronically in an open technical way can also facilitate internal and external validation of eCPRs using larger samples of EHR patients as a trigger for execution. As stated at the outset, impact analysis of CPRs was not considered in this work and the platform would need to be developed further to address the specific requirements for carrying out successful impact analysis of CPRs. The implementation aims for this research question in table 6-3 were achieved:

Table 6-3 Implementation aims of research question 3

<table>
<thead>
<tr>
<th>Implementation aim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demonstrate the viability of service-based provision of electronic diagnostic evidence through the construction of prototype implementations of the theoretical eCPR framework</td>
</tr>
<tr>
<td>Demonstrate the application of this technology to provide decision support using third party tools in practice</td>
</tr>
</tbody>
</table>

6.2 Delivery of Research Question Conceptual Aims

A number of conceptual conclusions can be made based on this research. The process as described for construction of data mined eCPRs was implemented as a two-step process. The definition of an eCPR model for representation of data is not enough on its own to support an eCPR development lifecycle. A process for evidence discovery is also needed using a more general diagnostic
model of evidence that enables us to describe and quantify more generally what we know about the association of diagnostic cues supporting any given diagnosis. From that general model we can construct eCPRs selecting subsets of ‘strong’ quantified diagnostic cues that can be represented in a more structured form of eCPR evidence.

I conclude that the process of eCPR discovery is a specific case of electronic diagnostic evidence discovery more generally and should not stand alone from it. The eCPR can be considered to be simply one defined ‘view’ of the evidence in a well structured format that is suited to a particular clinical purpose. This conclusion is consistent with the idea of eCPR knowledge being part of a broader epidemiological context as was discussed in the clinical literature review.

Other ‘views’ of the general evidence may also exist and could be constructed to present the general evidence in a more suitable format to support other clinical purposes such as on-line triage documents or general diagnostic guidelines for example. In effect a mapping is taking place of general evidence to a view of that evidence structured for a specific clinical purpose. The eCPR rule-based ‘view’ of evidence provides a useful tool for defining diagnostic rules that can be implemented as part of EHR systems.

Two important conceptual aims, as shown in table 6-4, were defined for this research in addition to the implementation aims previously discussed.
Table 6-4 Conceptual aims of research question 1

<table>
<thead>
<tr>
<th>Conceptual aim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop a theoretical framework describing a dynamic eCPR development and dissemination process, consistent with the wider goals of the LHS that provides a vocabulary to discuss and develop a technology driven vision for eCPR development.</td>
</tr>
<tr>
<td>Describe more ambitious diagnostic strategies using decision support tools that address the limitations of traditional CPR dissemination to facilitate wider clinical acceptance of CPRs</td>
</tr>
</tbody>
</table>

The conceptual framework for eCPRs was described in section 4 and the diagnostic strategy to make use of eCPRs for decision support was demonstrated in section 5.9. The eCPR maturity model as described is not a panacea to address all of the previously identified problems relating to the traditional CPR lifecycle. It has however been shown to provide a suitable framework to guide development of a real working implementation to support eCPR development as applied in section 5.

The eCPR framework provides a working ‘vocabulary’ to discuss and describe the components necessary to implement diagnostic decision support strategies. It provides a useful conceptual framework, consistent with the wider goals of the LHS, to describe the core functionality support required to make use of existing repositories of electronic patient data to implement eCPRs in practice. I argue that such a conceptual framework is necessary in order to develop more ambitious diagnostic strategies discussed in section 5.10.2 that address limitations of the traditional CPR development lifecycle. These diagnostic strategies are necessary to clearly define the role of eCPRs and how they fit into clinician workflows at point of care more generally. If we want clinicians to use decision support tools as a complement to their own professional judgment, the recommendations from such decision support tools should be targeted at
the right individuals and workflow rather than frustrate them with alerts and pop-ups that may not be relevant or necessary.

The conceptual aims of this research are presented with a view to promoting wider implementation and acceptance of the benefits that intelligent electronic use of eCPRs can provide. This research can encourage the wider use and acceptance of CPRs by clinicians in three important ways:

- by making eCPRs searchable and more contextually accessible than literature equivalents
- through support for provision of computable versioned rules derived from data mined sources of primary care data that are more sensitive to different patient populations
- through provision of eCPRs that can be embedded in decision support tools linked to EHRs workflows that facilitate easier use and execution.

6.3 Limitations of this Work

A number of limitations of this work need to be discussed relating to the underlying EHR data used as a basis for deriving and disseminating computable clinical evidence. These limitations are particularly related to the data mining approach and assumptions relating to clinical scenario selection and underlying EHR data quality that were made explicit in section 5.1. These assumptions were necessary to provide a manageable scope for this research work, concentrating on evidence representation and dissemination rather than underlying data gathering. A number of barriers to ‘true’ clinical learning will be discussed that focus on the lack of flexibility built into the underlying clinical terminologies used in structured data.

6.3.1 Bias and Error in Underlying Data Capture and Coding

The use of EHR data as provided for research purposes is not without problems. Problems can arise around data quality that inadvertently introduces bias or error when analysed. These data quality issues can be related to:

- missing data
• erroneous data
• un-interpretable data
• inconsistencies across data providers
• non-coded textual data (259).

For scoping purposes this research has made an underlying assumption that the EHR data provided is provided ‘as-is’ and was of sufficiently ‘good quality’ without explicitly exploring what that means, or doing any detailed investigation based on good practice to establish it (260). This was done on the basis that the data had already been previously used to do successful data mining using manual methods rather than ICT data mining methods (185, 261). A data mining approach at the scale of volume envisaged by the LHS would ideally ensure that quality procedures are followed that ensure consistency of coding of patient data across distributed locations. This would allow for meaningful comparisons and aggregation of that underlying data without risk that coding schemes are used differently in different local contexts.

The TRANSHIS EHR data used in this research was provided from across a number of different countries. The fact that the same EHR and coding scheme was used to capture that data does not necessarily guarantee comparable or good quality data in itself. In practice to guarantee good data mining results on aggregated data across institutional or jurisdiction boundaries we need to ask some relevant questions:

• is there the same consistency and a defined procedure and training of coding practice across different institutions or jurisdictions?
• at what timing points have final diagnostic codes been assigned to each recorded patient case and was a suitable diagnostic gold standard applied for confirmation? This has implications for whether a diagnostic label is actually accurate and was applied as a ‘working diagnosis’ in primary care or perhaps entered after confirmation through secondary care further testing which provides a different picture of the data
• are all patient cases over a defined time period fully recorded or are only subsets made available for privacy reasons for example?
• is there any required patient data that has not been recorded and is missing that would indicate a poor quality dataset generally?

The results of the data mining process on our data discussed in section 5.8. demonstrated consistency across different populations and the results agreed well with existing clinical literature and guidelines. This gives some reassurance that the quality of data coding was ‘good’ and consistent but it cannot be stated definitively. The differences in LRs shown between the Netherlands and Malta populations are certainly a function of the larger population available in the Netherlands for analysis but it may also be that different coding practices in each jurisdiction have introduced bias to the underlying data. Ideally a full assessment of data quality and coding practices relating to the underlying data sources used would be done but this was deemed out of scope for this work.

In the absence of such a definitive data quality assessment it is impossible to say that the eCPR construction process and the data derived eCPRs produced from it, are therefore free from underlying data bias. As I discussed in the literature review, heuristic approaches to clinical decision making are subject to various forms of cognitive bias including confirmation bias where the clinician pursues an initial diagnostic hypothesis even in the presence of subsequent contradictory evidence (44). These forms of bias can still manifest themselves in EHR coded data and associated data mining techniques (171, 262). The timing points of when data is entered into the EHR is not always obvious and may introduce bias in EHR datasets (263). From an epidemiological point of view confirmation bias may manifest itself in coded diagnostic evidence that is incomplete and does not capture a complete picture of the diagnostic workup constituting an error of data omission (264). This evidence as such may also constitute a temporal issue and relate to the diagnostic evidence that has been subsequently entered at a later date to confirm an existing diagnostic hypothesis. In such a case therefore there is no certainty that our data derived eCPRs reflect all the possible diagnostic associations that could be discovered through analysis of an EHR that is known to capture all consultation diagnostic cues at the point of care.
The TRANSFoRm project has highlighted the importance of this issue and suggested a way of tackling this problem. This was done by developing a tool to present diagnostic evidence at an early stage of the consultation with supporting prompted evidence supporting a range of differential diagnoses to consider at the outset. The tool importantly supported integration with the EHR to allow saving of a full set of consultation evidence back to the EHR to support subsequent data mining efforts. The TRANSFoRm evaluation of the diagnostic tool it was found that ‘using the CDSS prototype resulted in, on average, 12 times more data coded into the EHR during the consultation’ (258). The temporal aspects of evidence recording bias were also highlighted where most clinicians using the EHR only ‘recorded information only at the end of the consultation, after the patient had left the room’ (258).

6.3.2 Granularity of Coding Scheme Used

A crucial issue that arose during this research was the importance of giving due consideration to the granularity and expressiveness of the underlying coding schemes used for data mining aggregated patient data.

The data mined evidence from TRANSHIS was based on coarsely granular ICPC2 EHR data. ICPC2 is widely used in primary care and is well suited to capturing high level consultation data. The third party developed TRANSFoRm decision support tool was linked to an EHR that implemented a coding scheme on a much finer granularity in the form of NHS read codes. These codes capture consultation data at a much lower level of granularity.

Vocabulary binding of clinical evidence has been done using manual configuration of the evidence ontology with NHS read codes which is not sustainable in the longer term. Future work is needed to provide integration with a clinical vocabulary service application. This could allow for a vocabulary enrichment process by using terminology bindings manually defined in the ontology to identify and populate mappings to other terminologies from a central vocabulary service that can be dynamically populated into the ontology content.
Matching the ICPC2 derived diagnostic evidence by triggers from NHS coded patient EHR data was therefore problematic without some sort of terminology mapping to link the two terminology sets. In effect there was a mismatch in the granularity of coding schemes used at both ends of the data mining evidence generation and dissemination pipeline of the LHS infrastructure. It could therefore be argued that ICPC2 was not a good choice for data mining as the high level of granularity meant it was not expressive enough from an epidemiological point of view to express diagnostic scenarios supporting complex diagnostic decision-making. Manually curated diagnostic evidence coded using NHS read codes and based on clinical review of best practice guidelines was therefore necessary to fill that gap to construct clinical scenarios with sufficient levels of clinical detail to demonstrate and evaluate the utility of the decision support tool itself as part of a controlled trial.

The alternative to using ICPC2 data for data-mining also has drawbacks however. Because ICPC2 is at a high level of granularity it was possible to aggregate and analyse large numbers of cases related to each combination of ICPC2 codes, thus resulting in stronger associations in the form of larger likelihood ratio values and tighter confidence intervals. If we use a finer level of granularity we then split a number of cases amongst lower level granular codes. This will reduce the support value (number of cases) for each association rule for each code. What was previously X support cases associated with one high level granularity code is now X cases divided and distributed among a number of lower level codes. This reduces the number of potential cases available for analysis which in turn will reduce the strength of the evidence that we can generate in the form of larger LR with wider confidence intervals (as was seen in the Maltese data where the population was smaller). There is therefore a trade-off to be considered between the level of granularity that we code at and the desired strength of evidence.

In conclusion it can be stated that careful consideration and selection is needed of the underlying EHR structure and coding schemes used for knowledge learning. The work would be significantly improved with the ability to integrate
the use of a vocabulary service to dynamically map at runtime to target coding schemes used by EHRs where the knowledge will be deployed.

6.3.3 Limited Literature Based Evaluation of the Data Mined Evidence

The evaluation of the data mined evidence supporting a diagnosis of Urinary Tract Infection and Pyelonephritis as presented in section 5.8.3 has focussed on a comparison of that evidence with a selection of literature based evidence in the form of clinical reviews and international guidelines.

Ideally an alternative and more robust mechanism would be followed according to the traditional CPR development lifecycle, validating the performance of that evidence to diagnose Urinary Tract Infection and Pyelonephritis against a second ‘testing’ dataset of EHR data, separate from the original derivation or ‘training’ dataset population, obtained from the TRANSHIS database. The second testing dataset would contain a representative sample of patients from TRANSHIS with any recorded diagnosis.

The data derived eCPRs as described in section 5.8.9 could be applied to the testing dataset to determine the sensitivity of the eCPRs by comparing those identified by the eCPR as ‘high risk’ for those conditions with those having an actual recorded diagnosis of Urinary Tract Infection or Pyelonephritis. The specificity of the eCPRs could also be assessed by comparing those assessed as ‘low risk’ with the final recorded diagnosis. This would in effect constitute a ‘narrow’ validation of the derived eCPRs by applying it to a different dataset to the original derivation dataset from the same population. This approach is not without problems as it is subject to the same data quality and bias considerations that were discussed previously in section 6.3.1.

A practical consideration was the lack of availability of a second ‘testing dataset’. The original TRANSHIS dataset could have been split into two datasets but this would have decreased the number of records available for derivation and the strength of the derived clinical evidence. As derivation was the focus of this work I wanted to maximise the number of records available to
derive quantified evidence. On that basis a comparison with other third party literature based sources of clinical evidence was deemed appropriate.

6.3.4 A Limited eCPR Learning Cycle Without a Feedback Loop

In a true LHS the consultation data captured as part of the decision support tool would be fed back into the original derivation data set to enable further cycles of data mining to enable eCPR performance characteristics to be developed on an iterative basis. This work has demonstrated the feasibility of deriving quantified eCPR knowledge through a single cycle of data mining but did not implement a feedback loop to see how the derived eCPRs might subsequently change as new patient data is gathered.

As previously discussed, the data derived eCPRs in this research were not subject to a formal process of internal or external validation and are therefore considered to be at a low level of maturity. The experience with traditional CPRs has demonstrated that lack of CPR maturity is a barrier to CPR uptake in practice and only mature CPRs that have become embedded in clinical guidelines are widely used in GP practice (34, 88). The architecture as implemented does not distinguish between eCPRs that may be at different levels of CPR lifecycle maturity (derived, internal validation, external validation, impact analysis).

More work is therefore needed to develop what could be described as separate ‘test’ or ‘production’ environments within the overall ICT architecture. In practice this is easily enough achieved by using two separate backend evidence stores and making the evidence service configurable to point to either environment. The end user decision support tools could then be configured to use evidence in a ‘test’ capacity that could be applied to test extracts of EHR data for the purposes of performing eCPR validation studies. This mode of operation combined with a feedback loop as discussed could facilitate a full learning eCPR lifecycle.

The focus of this work was on the development of the backend evidence derivation and not on the development of the decision support tool itself which
did not implement such a feedback mechanism. Further work is necessary to
demonstrate this cyclic approach to eCPR derivation and validation in practice.

6.3.5 Constraints on Underlying Terminology Unable to Facilitate
Epidemiological ‘Unknowns’
Coding schemes by their nature describe semantically well defined concepts to
represent ‘what is clinically known’ within a defined clinical domain. They do not
describe conceptual unknowns which presents a significant problem for real
clinical evidence learning. If the goal of a learning health system is discover new
clinical knowledge, it is arguable that the data mining methods presented in this
research cannot achieve this using an already defined clinical terminology.
Realistically this approach will only allow us to quantify and confirm or disprove
what we already consider to be knowledge as defined by existing clinical coding
schemes.

The application of more ‘fuzzy’ data mining techniques such as latent class
analysis, for example, may be needed to discover and represent ‘clinical
unknowns’ within our models that cannot be captured using a predefined set of
clinical codes alone. Rindskopf for example describes how latent class models
of diagnosis that can be built even in the absence on a confirmed but possibly
unreliable coded diagnosis by defining two latent classes to represent patients
that do or do not have a particular clinical condition (265). A similar approach
could be used to define latent ‘symptom’ classes to highlight relationships of
these symptom ‘unknowns’ in relation to diagnosis of specific conditions. This
could provide insights to identify what they actually clinically are resulting in
generation of new actual clinical knowledge.

6.3.6 Clinical Scenario Difficulty and Selection for Primary Care Settings
The clinical scenarios selected to demonstrate the data mining approach
deliberately focus on what can be considered a well understood and common
presenting complaint in primary care practice. As such this provides large
volumes of coded data to investigate those particular problems. Tackling more
complex diagnoses in primary care is more difficult.
The probability of a family practitioner diagnosing a case of cancer in family practice is low. Cancer is by its very nature a more complex disease to diagnose. A lack of coded ‘confirmed’ evidence cases in primary care combined with a lack of granularity provided by ICPC2 to represent diagnostic evidence for cancer is problematic. This results in a diagnostic use-case that is not easily data mined or comparable with suitable clinical literature at the correct level of detail.

Another improvement that might be made is that data mining has been carried out to produce association rules based on the first encounter of new episode of care. Subsequent development could look at doing analysis at further encounters and timing points in addition to the first encounter of an episode of care. This should allow for better analysis of more complex conditions which may take more than one encounter to formulate a definitive diagnosis. This could also allow for time-based concepts to be added to the ontology by comparing rules generated for conditions at different points in time.

6.4 Discussion and Future Work

I have concluded that generation of meaningful clinical evidence for diagnosis is possible using electronic sources of primary care data. A number of important topics for discussion that will ultimately determine the success or failure of developing the learning decision support tools come out of this work and need to be considered in the future.

Further policy work is needed to ensure that these tools can make use of the growing bodies of electronic health record data to support derivation and dissemination of diagnostic evidence in practice, supporting clinicians and driving real improvements in patient safety. Assuming that clinical evidence can be disseminated for use by clinicians in diagnostic assessments with real patients, we must then consider new models of responsibility combined with rigorous quality assurance processes for generating and managing that clinical evidence. This involves two main areas of focus that are non-technical challenges and are arguably more difficult to address:
• quality assurance procedures for software development, software updates and evidence updates – treating the DSS as a ‘medical device’
• medico-legal issues that patients and practitioners who will use these tools need to be assured about

We also need to address what happens if things go wrong. Is there a governance to fully investigate and rectify problems? This needs to address what happens when that evidence results in adverse events or recommendations for patients.

6.4.1 Implications of Medical Device Directives for this Work

In the literature review I highlighted the increasing importance of certification and regulation that attempts to guarantee high quality software development where software is considered to be ‘medical device’. Software that provides diagnostic decision support tools certainly fall within this remit and although this is a research work, the practical implications for developing real-world implementations must be given due consideration.

The practical need for more research in this area has been starkly highlighted by recently reported news events in the UK (266). As previously described, the QRISK 2 score is a validated, accepted and widely used decision support aid for predicting the cardiovascular risk in patients in the UK (127). It has been successfully implemented as a standalone online electronic tool and has subsequently been integrated with third party general practice electronic health record systems via a parameterised programmatic interface (221).

In the case of an EHR integrated QIRSK2 rule, an interface between the EHR system and the QIRSK2 calculator allows the EHR to trigger the rule from within the EHR. The EHR can contextually trigger the rule for a chosen patient record, sending the required data for that patient via a programmatic interface to the rule algorithm, receiving back a risk score interpretation for that patient. On the basis of a high risk score being returned, a clinical decision may be recommended by the GP to put the patient on a course of statins.
In one particular implementation of the QRISK2 score with a widely used third party general practice EHR system something appears to have gone wrong. It has now retrospectively come to light that the reported cardiovascular risk has been overstated in some patients resulting in the possibility that they may have been incorrectly advised to take statin medication to lower that risk. This has resulted in a full investigation by the Medical and Health Products Regulatory Agency (MHRA) in the UK.

The initial suspicion has focussed on the interaction and submission of patient data between the EHR and the QRISK2 calculator via the interface, rather than the implementation of the QRISK2 tool itself. This has resulted in the need for the EHR vendor in question to identify GP practices to notify potentially impacted patients and to review their cardiovascular risk again (266). This raises significant questions relating to the implementation of more distributed forms of decision support that might be classified as medical devices as the actual QRISK2 algorithm was correct but the actual implementation in the EHR system has gone wrong.

The development of the concept of ‘software as a medical device’ originated when software was largely implemented as single ‘app’ tools rather than web based services or components. The situation regarding a distributed infrastructure such as the LHS is less clear however: where does the medical device start and where does the medical device end? In the context of this work you could argue that the decision support interface, the evidence service and the data mining module could all be considered medical devices separately in their own right. In the literature I discussed the complexities of the certification process that is implemented based on deployment of the software in each country and jurisdiction. Should components of the LHS be certified individually or as a single system and what is the regulatory situation if different components are deployed across different legal jurisdictions as was the case in the TRANSFoRm project?

I suggest that the introduction of the concept of a ‘medical device chain of dependency’ needs to be considered in this context to identify distributed components that are critical to the safe clinical application of end-user tools.
being used directly with patients. Certification of the end-user tools should only be granted when certification of the underlying dependent components has also been granted. If the dependent components are distributed in different jurisdictions it may be that they also need certification in the jurisdiction within which the end-user tool is deployed for use with patients.

A further problematic issue to consider in this context is the use of open-source tools. A design objective for this research work was to maximise use of open-source technical tools and standards (KNIME, Protégé, Sesame, Tomcat). This gives assurance that the coding used to implement diagnostic algorithms is openly available and open to scrutiny which can be seen as hugely beneficial compared to ‘black-box’ proprietary solutions where the decision making process is not open to review. Medical device legislation suggests that guarantees of quality assurance will also be required for these open-source software solutions which are outside the direct control of developers that build upon those tools. This requirement has the potential to severely restrict the implementation choices available to those developers who wish to use more openly available technical tools and standards.

Medical device legislation should ensure that all required components of such systems are implemented to high standards and that the legislation does not allow for ‘weak points’ that may cause adverse events with patients. Some sort of ‘registry of approved open-source technologies for medical devices’ would also be of benefit to developers in making decisions regarding implementation tool choices. Ensuring software quality is implemented to an appropriately high level is an absolute pre-requisite for developing trust in decision support tools and the concept of the LHS.

6.4.2 Developing Trust in the Learning Health System

The wider uptake of learning decision support systems with the potential for generation of clinical evidence is dependent on the generation of a critical mass of trust amongst the clinical community who are the potential target users for these systems. Patients also need to trust that these systems are a benefit to clinicians and do not threaten standards of care. If that critical mass is not
achieved then learning diagnostic decision support systems will not be more widely trusted and accepted.

These concerns were articulated succinctly by Dr. Gerald P. Corcoran in a discussion forum relating to the role of ICT in diagnostic decision support when he commented that ‘anyone is entitled to seek advice about medical symptoms from anyone. Often a neighbour, the butcher, the hairdresser all have input, but the seeker weighs their advice appropriately’ but ‘Dr. Google is nowhere to be found when responsibility needs to be assigned’ (267).

From a clinician’s point of view, the development of trust is therefore closely tied to the concept of ultimate responsibility and the quality of the underlying evidence. The importance of curation of generated knowledge has been as an essential part of the LHS infrastructure in this research work but curation of evidence should also imply taking responsibility for it. In the literature review we have seen that adverse events will always take place within our health systems. If we develop new ways to derive clinical evidence electronically then it is my opinion that we must also be explicit in assigning responsibility for that use of that evidence. This also implies developing governance procedures for tracking how that evidence is used and any adverse events that may result from its use in clinical practice.

To move the LHS beyond the realms of research projects means making the evidence generation elements of the LHS sustainable in the long term. This will require establishment of appropriate organisational, national or professional bodies who are willing to take on the responsibility for ensuring standards of evidence generation inspire trust and that on-going evidence is maintained. National research networks would appear to be the obvious candidates for such a role. This may imply that those bodies should also accept medico-legal responsibility for the generated evidence.

The medical device regulation process may provide another more market-driven alternative to act on a trust basis with appropriate disclaimers relating to ultimate responsibility. Some would argue that decision support tools should be no different in their treatment to how we already use the internet for seeking
clinical guidance; evidence is presented ‘as-is’ and the consumer makes a judgement on the trustworthiness of the source of that evidence. The hope would be that if clinical evidence is made transparent and openly available for scrutiny along with demonstrable labelling relating to guarantees of quality, clinicians will drive uptake of these tools while still assuming ultimate responsibility for the medical decisions they make when using them. If those tools are deemed to clinically useful then they will succeed, otherwise they will fail.

6.4.3 **Structured vs. Unstructured Clinical Evidence Discovery**

The previous concerns expressed about the limitations of clinical terminologies to allow for true clinical evidence learning lead to a wider discussion of the type of data mining approaches we should employ for clinical applications. We have previously discussed the critical characteristics of interoperability and related those to our conceptual model for implementation. A critical aspect of interoperability is the need to specify meaning of data or ‘semantics’. As we have seen, in the case of health data this relates largely to the adoption of suitable terminology or vocabulary services to allow binding of specific vocabulary terms that apply semantic meaning to explicit model terms. The meaning of these terms is restrictive but explicitly defined to allow aggregation and analysis of data with some degree of confidence that the aggregated patient cases are grouped appropriately.

The alternative data mining approach is the increasingly investigated field of research relating to data mining using ‘free-text’ or unstructured data such as social networking posts. Typically this pre-processing of text to allow extraction of key ‘terms’ from textual data which are then tagged or mapped to an agreed set of terms describing the knowledge domain of interest (164, 174). The processed text can then be aggregated and analysed using data mining techniques to enable knowledge discovery. The benefits of this approach allow those contributing data to do so using natural language descriptions rather than a set of ‘codes’. The application of this approach is therefore is suitable to less ‘formal’ communication mechanisms such as aggregations and analysis of social media message content.
In the specific context of discovering diagnostic knowledge it appears that any such approach presents its own problems and has potential to introduce ambiguity associated with the aggregation of patient data. The need to define an internal list of terms that define the knowledge domain to which free-text can be mapped implies that the approach is still comparable and has similar limitations to a structured approach. There is more potential however for ambiguity since standardised clinical vocabularies may not be used and a mapping process for free-text terms has also to be defined.

I contend that despite their limitations the use of agreed clinical terminologies using a structured data mining approach is a critical aspect for the successful implementation of the Learning Health System. The need for unambiguous capture of patient data in EHRs does not in my opinion lend itself to unstructured approaches using textual notes only. Only through a strict definition of semantic meaning can we ensure that we are aggregating data of similar semantic types with a view to detecting trends and patterns in data that allow meaningful quantification of relationships. What is needed is more research to build more flexibility into vocabularies and terminologies that allows the same terminology or vocabulary to be ‘viewed’ or applied in more than one clinical setting or context. For example a single vocabulary might provide views or subsets that are present terms at a suitable level of granularity that are suitable for primary care data capture, secondary care data capture or epidemiological representation.

### 6.5 Contributions to Knowledge

The contribution of this work should be seen in two particular contexts that will now be considered. The specific aims of this research addressed specific gaps relating to CPR theory itself. A wider context allows us to discuss the contribution of this work in terms of broader development of understanding relating to threats to patient safety posed by diagnostic error in practice.

#### 6.5.1 Contribution to CPR Theory and Knowledge

The first narrower context is the demonstration of the feasibility of an alternative EHR data-driven process to derive and disseminate eCPRs that addresses a number of inherent problems found in the traditional CPR development lifecycle.
The work has placed CPR development in the context of a learning health system and demonstrated the implementation of a conceptual framework for eCPR development in practice. It is hoped that other researchers with an interest in CPRs can take the conceptual framework developed here, develop it further for their own needs, and use the implementation approach described here to suggest or guide research work in their own particular environments using electronic health data.

At the outset of this work a number of gaps in knowledge were identified specifically relating to current practice regarding derivation and dissemination of CPRs. For clarity these are collected in table 6-5.

Table 6-5 Gaps in knowledge regarding use of CPRs in practice

<table>
<thead>
<tr>
<th>Knowledge Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong></td>
</tr>
<tr>
<td><strong>2</strong></td>
</tr>
<tr>
<td><strong>3</strong></td>
</tr>
<tr>
<td><strong>4</strong></td>
</tr>
<tr>
<td><strong>5</strong></td>
</tr>
</tbody>
</table>

This development of published models of eCPRs described in sections 5.4, 5.5 and 5.6 has specifically addressed knowledge gaps 2 and 3 (227, 228, 235). These sections describe eCPRs that are computable, generalisable, interoperable using open standards, and searchable. A published and working
demonstration of how these models can be consumed by decision support tools in practice was demonstrated through the development of a diagnostic workflow process and tool integration as described in section 5.9 (229).

A published process for the development of clinically meaningful eCPRs from data mining applied to electronic sources was shown to be achievable in section 5.8 (172, 184, 245). This directly addresses knowledge gap 4, subject to the previously acknowledged limitations related to the underlying EHR data that was used.

Knowledge gap 5 was addressed in section 4 and describes a published conceptual framework for an LHS infrastructure for eCPR development (204). This framework provides the research community with an approach and understanding of the core elements that are needed to be put in place an electronic diagnostic evidence generation process to support eCPR development. The framework specifically addresses both derivation and dissemination of eCPRs.

The collective pieces of this research work as a whole demonstrate how knowledge gap 1 could be addressed. More work however would be needed to take eCPRs and to actually validate these in practice using the tools that have been developed here. It remains a question yet to be answered whether that can definitively drive greater awareness and uptake of CPRs in clinical practice.

6.5.2 Diagnostic Error in Family Practice

The wider context and contribution of this work though can be seen in light of recent calls for more research to address diagnostic error in practice, now recognised as an issue that needs to be meaningfully addressed as part of a broader patient safety reporting culture (52, 54). As ever, the growing influence and potential for ICT is recognised in these calls as one stakeholder in a larger process.

The Committee on Diagnostic Error in Health at the Institute of Medicine in Washington made a number of important recommendations in this area in 2015 (54). Specifically they recommended the development of ‘a coordinated research agenda on the diagnostic process and diagnostic errors by the end of
2016’. Recommendations were made relating to goal 3 and 8 of the report which are highly relevant to this work:

- **‘Goal 3: Ensure that health information technologies support patients and health care professionals in the diagnostic process**
  - Recommendation 3a: Health IT vendors and the Office of the National Coordinator for Health Information Technology (ONC) should work together with users to ensure that health IT used in the diagnostic process demonstrates usability, incorporates human factors knowledge, integrates measurement capability, fits well within clinical workflow, provides clinical decision support, and facilitates the timely flow of information among patients and health care professionals involved in the diagnostic process.
  - Recommendation 3b: ONC should require health IT vendors to meet standards for interoperability among different health IT systems to support effective, efficient, and structured flow of patient information across care settings to facilitate the diagnostic process by 2018.

- **‘Goal 8: Provide dedicated funding for research on the diagnostic process and diagnostic errors**
  - Recommendation 8a: Federal agencies, including HHS, the U.S. Department of Veterans Affairs, and the U. S. Department of Defense, should:
    - **Develop a coordinated research agenda on the diagnostic process and diagnostic errors by the end of 2016.**
    - Commit dedicated funding to implementing this research agenda.
  - Recommendation 8b: The federal government should pursue and encourage opportunities for public–private partnerships among a broad range of stakeholders, such as the Patient-Centred Outcomes Research Institute, foundations, the diagnostic testing and health IT industries, health care organizations, and
professional liability insurers to **support research on the diagnostic process and diagnostic errors**.' (54).

These recommendations make reference to the inclusion of ‘a broad range of stakeholders including health IT. The research approach to diagnostic decision support described in this work can be seen as highly relevant with respect to making a practical contribution to addressing those recommendations. This research is also consistent with calls highlighting the important role the EHR can play as a trigger for personalised healthcare delivery (268). This research has demonstrated that the goals of the Learning Health System can be closely aligned with the development of next generation diagnostic decision support tools in order to address these challenges.

A growing awareness of patient safety issues and the need for national incident reporting is also developing in Ireland. Recently proposed bills along with patient safety recommendations suggest a strategic role for ICT in promoting patient safety. The recently published Health Information Quality Authority report on the coordination of patient safety intelligence in Ireland stated in recommendation 4 that there was an urgent need to establish:

- **‘an effective information communications technology infrastructure, supported by standards for interoperability should be put in place to support the coordination of patient safety intelligence’** (52).

The development of tools based on open standards that promote more proactive approaches to developing a patient safety culture through evidence based practice, as proposed in this research work, can be seen as a positive contribution consistent with this longer term goal.

**6.6 Generalisability of Research and Future Work**

A primary focus for this work was the development and application of ICT solutions based on open standards without use of proprietary technology. The client decision support system application shown here that consumes the service was accessed and integrated with a primary care EHR system. However the evidence models developed here are generalisable so that they could be populated to include additional clinical content beyond the presenting...
clinical conditions and eCPRs presented as part of this work, or to other clinical contexts beyond primary care. As such they are suitable for representing any eCPR that is based on scoring the presence or absence of presenting patient symptoms. A number of other eCPRs relating to difference conditions have already been represented by these models including the Keith Edwards score for childhood tuberculosis (269), the acute bronchitis severity score (270) and the Finnish Diabetes Risk Score (271). The models and dissemination mechanism are therefore not tied to any particular disease or clinical context.

Because of the web service based design of the knowledge base and the openly available access to it, other areas of potential use can be suggested that are not directly related to clinical practitioners. The development of patient accessible ICT solutions has been a huge research growth area in answer to criticism that health interventions have focussed too much on clinician controlled data in the form of the EHR. There have been calls to support development of increased public health literacy through the use of a ‘public API for health data that can catalyze the development of an ecosystem of apps, for both clinicians and patients, that run on health data’ (272).

The open models and service oriented architecture that we have developed are flexible enough to allow any third party tool using open standards to ask diagnostic questions using standardised web queries. The fact that the knowledge base is accessed as a separate service means the architecture can support additional functions to empower patient-oriented approaches to healthcare in including self-triage or safety-netting related to serious conditions for which suitable CPR evidence is available. This could be done through provision of evidence based content accessible using smartphone or tablet based third party clients for use by the general public.

Development of new models of evidence dissemination that directly engage and educate the general public in appropriate self-management symptoms can have benefits for efficient use of existing health system resources (273)(274). Health system resources can be more efficiently used through empowering patients to self-manage illness where appropriate without the need for unnecessary
referrals, along with the education and promotion of appropriate antibiotic use only where necessary (275, 276).

The research work here also has potential to demonstrate health service impacts relating to evidence based practice. It demonstrates a general framework for eCPRs that enables the use of research-informed interventions in the health system using models based on the Learning Health System. This can facilitate faster translational research of evidence based guidelines using open ICT architectures and smartphones as an alternative to literature based guideline dissemination. The data mining approach described supports the gathering of large volumes of patient reported data required for direct patient based observational epidemiology.

6.7 Is Ireland Ready for the Learning Health System?

As a final discussion point it is desirable to provide some context on where Ireland sits in relation to broader use of ICT in a health context and to issue a ‘call to arms’ relating to our ICT health infrastructure. The fact that we have such a large presence in this country of ICT and healthcare multi-nationals should give us huge potential to take an active part in this emerging research area. Whilst acknowledging that the LHS can be implemented at organisational or regional scales, more ambitious visions see the LHS at national or even global scales (159). Ireland is currently in the process of developing its own health ICT infrastructure and as such may have an opportunity to develop it with a longer term national LHS in mind that supports national research. The question as to whether or not Ireland is ready for a national LHS is deliberately provocative and might be considered inappropriate on the basis that the LHS is still very much in the realm of research.

It should be clear based on what we have discussed that very few countries could claim to be ready for a national LHS. It should be stated however that other countries have made significantly more progress in putting in place the core prerequisites that could make it happen, particularly in the areas of developing widely used EHRs and associated research networks. The area of ICT investment in healthcare has been sorely neglected in Ireland and we have
much work to do to put the basic core ICT elements in place before even contemplating the construction of anything like a learning health infrastructure.

The good news is that the landscape in Ireland has finally changed over the course of the development of this research and there are now reasons for some optimism. Some excellent work is now being done with the establishment of governance structures in the form of the recently appointed Chief Information Officer (CIO) and Health Ireland group, along with policy development in the form of the eHealth and Knowledge and Information strategies (121, 277). This is driving the required engagement with a broad range of interested stakeholders. We are now making the required policy decisions and recommendations from the Health Information Quality Authority that will provide the core building blocks for future work such as the provision of patient health identifiers and identification of best practice standards (52, 278). If implemented correctly, these fundamental policy decisions can put in place the prerequisite building blocks necessary for the future delivery of an LHS like infrastructure in the form of national research networks.

The learning health system presents us with a very challenging multi-disciplinary vision that critically includes the ability to carry out large-scale electronic population based research. We need to include the role of electronic research networks in our national discussions and start planning for the establishment of those networks across clinical domains for the purposes of population based research. This requires explicit policies that are compatible with the relevant EU legislation promoting responsible secondary use of electronic patient data whilst guaranteeing patient privacy. This can facilitate an agile electronic research and an evidence generation process that can benefit and inform national health policy makers, clinicians and patients.

6.8 Summary and Conclusion

Evidence-based medicine has long been advocated as one way of encouraging clinical reasoning that is based on a more rigorous and systematic approach, without reliance on more heuristic methods of decision making. The effective practice of evidence based medicine will never be fully delivered unless better
ways can be found to disseminate clinical findings generated from research environments into front line clinical practice.

Many have also argued that the value of evidence based medicine should not be overstated and incorrectly presented as a 'silver bullet' for addressing all patient safety issues. They argue that it results in volumes of top-down enforced guidelines that don’t put value on individual clinician judgement and quickly become out of step with the changing demographics of underlying patient populations (6, 59, 65-67). The problem may not be the practice of evidence based medicine itself but rather the current ‘means of production’ to support it and disseminate it.

The often politically talked about ‘knowledge economy’ is ‘an economy in which growth is dependent on the quantity, quality, and accessibility of the information available, rather than the means of production’ (279). This dictionary definition is interesting in that it relegates the important ‘means of production’ to an afterthought; it becomes secondary to the stated principle activity of producing large quantities of widely accessible information. This unbalanced emphasis implies that the means of production is not crucial in itself to discover ‘information’ of sufficient ‘quality’ and ‘quantity’ to make high quality decisions.

When we consider the clinical knowledge necessary to inform decisions relating to population health policy and individual citizens health, the ‘means of production’ merits far greater consideration and emphasis. Our goal should be to implement a process that comes with the appropriate tools to support rapid derivation of high-quality clinical evidence for informed decision-making within our health systems. The growing body of available EHR data means the application of ICT can provide an alternative ‘means of production’ of clinical evidence. It is also important to state that these alternative means of production will still need curation and review by clinical professionals. The traditional roles played by clinical professionals may need to change as part of a new process to support quicker knowledge production.

The generation and representation of actionable clinical knowledge from electronic sources of patient data is obviously a challenging task. The platform and associated models developed as part of this research have demonstrated
that it is feasible to generate and make openly available clinically meaningful knowledge that complements available literature based evidence for diagnostic purposes. The challenge is to identify suitable patient data sources in large enough volumes from which evidence of a sufficiently high level of confidence can be generated and represented. The open implementation and integration of evidence from empirically measured EHR sources described here is consistent with the complex goals of developing a learning healthcare system. The potential for using large collections of aggregated electronic patient data to support these goals appears underutilised at present and requires further research and development.

I have argued that the LHS is a suitable means for implementing such systems. Why is the LHS important? It is easy to dismiss the learning health system as an exercise rooted in the pursuit of academia. The fact that the European Union and organisations showing significant interest in this research come from a wide ranging spectrum beyond academia, including EHR vendors and pharmaceutical companies, would indicate otherwise. These groups are actively engaged and discussing the practical opportunities this represents right now. The technical solutions required are not incomparable to those already being used to support organisational business intelligence albeit that they do it at a much smaller scale.

The policy issues are arguably far more difficult to address but can be discussed and planned for now without the need for infrastructure development just yet. The secondary use and sharing of patient data is controversial as commercial interests are seen to increasingly drive the lucrative health research agenda both in the US and in Europe. In Ireland a lack of clinical research and an awareness of the benefits of secondary use of electronic data combined with the need to alleviate privacy concerns through establishment of patient consent agreements are issues to be addressed (280, 281).

The revelations and privacy issues arising from the Edward Snowden affair have understandably made citizens even more sceptical and governments more reluctant than ever to tackle this ‘hot potato’. Only through patient inclusion, education and informed policy formation can the benefits of ambitious efforts
such as the LHS be presented for what they are and not just another back door for ‘big-brother’ style interference. The danger is in choosing not to pursue an ambitious vision but instead letting commercial vested interests take sole control of this important research agenda; allowing them to ultimately shape the dialogue and policies, reap the benefits and relegate to a secondary concern the driving goal of improving clinical practice, patient safety and patient outcomes.

Whilst technical challenges also need to be addressed, a huge amount of research effort and debate has been expended over the types of clinical data models that we devise to represent clinical data. Whilst fully recognising the need for defined data representation standards to support system interoperability, I contend based on this research that the increasing availability of open source software combined with the ability to easily translate data from one data format to another, has meant efforts in this area are no longer the priority that they once were. As previously discussed, real problems lie with reconciling our semantic representations of that data and how that limits us in aggregating and making sense of the bigger picture. On that basis I would call for the focus of research effort to shift from system interoperability issues to semantic interoperability issues. If we do not get the semantics correct and build in more flexible mechanisms to view data in semantically different ways then I fear that the delivery of the LHS will be a failure.
7 References


11. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research


68. Bennett A. The history boys: a play: Macmillan; 2006.


120. eHealth Action Plan 2012-2020 - Innovative healthcare for the 21st century. Communication from the commission to the European parliament, the council, the European Economic and Social Committee and Committee of the Regions; 2012.


229. Corrigan D. An ontology driven clinical evidence service providing diagnostic decision support in family practice. AMIA Summit on Clinical Research Informatics; San Francisco, USA 2015.


Joint Workshop on Semantic Technologies Applied to Biomedical Informatics and Individualized Medicine; 2012; Boston, USA.


258. Final validation of prototype DSS: TRANSFoRm Project; [22/07/2016]. Available from: http://www.i-hd.eu/i-HD/assets/File/TRANSFoRm/D2_2%20Final%20validation%20of%20DS S%20prototype_TRANSFoRm.pdf.


263. McKillop M, Polubriaginof F, Weng C. Examination of Temporal ICD Coding Bias Related to Acute Diseases.


Appendix A – Search Criteria for Diagnostic Evidence Reviews of Literature

The search was carried out focussing on identifying clinical papers containing diagnostic evidence relating to the following categories:

- Demographic or lifestyle risk factors
- Physician observer diagnostic signs
- Patient reported symptoms
- Associated diagnostic Clinical Prediction Rules
- Prevalence information as available by demographic region

**Step 1**

An initial search was carried out for the target diagnosis on the HRB Centre for Primary Care Research database of clinical documents which was created to identify clinical papers containing clinical prediction rules or treatment algorithms from a selection of 28 primary care focussed medical journals. This database contains 70,000 clinical papers covering 1980-2008.

An example of the search string used for creating this database from PubMed for the period 2008 is:


**Step 2**

The following evidence based websites were searched for articles relating to the diagnosis:

- JAMA Evidence
- British Medical Journal Clinical Evidence
- The Cochrane Medical Reviews
- NHS Clinical Knowledge Summaries
- NICE Guidelines
- SIGN Guidelines

**Step 3**

A search of pubmed and embase. An example search string used was:

'tuberculosis'/de AND 'diagnosis':ab,ti AND ([article]/lim OR [review]/lim) AND [humans]/lim AND [english]/lim
References related to retrieved articles were systematically searched.
Appendix B – Example of SPARQL Query Execution Using Sesame API

public static EvidenceDiagnosis query(String rfe) throws OpenRDFException {
    // Create evidence diagnosis object
    EvidenceDiagnosis evdDiag = new EvidenceDiagnosis();

    // Create array for holding diagnosis list returned from SPARQL query
    ArrayList<EvidenceDiagnosis> nameList = new ArrayList<EvidenceDiagnosis>();

    // Create Sesame repository pointing to the clinical evidence
    // repository ID
    Repository myRepository = new HTTPRepository(Constants.sesameServer, Constants.repositoryID);
    try {
        // Initialise the Sesame repository connection
        myRepository.initialize();
    } catch (RepositoryException e1) {
        e1.printStackTrace();
    }
    try {
        // Establish the connection to the repository connection
        RepositoryConnection con = myRepository.getConnection();
try{

// Create SPARQL query string to retrieve differential diagnoses based on presenting RFE

String queryString = Constants.queryString + "SELECT DISTINCT ?anyDifferentialDiagnosis WHERE{evd:+ rfe + evd:hasDifferentialDiagnosis ?anyDifferentialDiagnosis.}";

System.out.println(queryString);

// Execute the SPARQL query against the evidence repository

TupleQuery tupleQuery = con.prepareTupleQuery(QueryLanguage.SPARQL, queryString);

//System.out.println(tupleQuery);
TupleQueryResult result = tupleQuery.evaluate();

//System.out.println(result);
System.out.println("Tuple Query result");

try {
    // Iterate through each differential diagnosis and extract diagnosis properties
    while(result.hasNext())
    {
    ...
    

    ...
{ 

    BindingSet bindingSet = result.next();
    Value anyDiffDiag =
    bindingSet.getValue("anyDifferentialDiagnosis");

    String diffDiagnosis = anyDiffDiag.stringValue();
    diffDiagnosis = diffDiagnosis.substring(66);
    evdDiag.setEvidenceDiagnosis(diffDiagnosis);

    // Add the diagnosis object to the evidence diagnosis list to be
    // returned
    nameList.add(evdDiag);

    }

} 

finally
 {
    // Close SPARQL result set
    result.close();
}

} 

finally
 {

    // Close Sesame connection
    con.close();
}

} 

catch(OpenRDFException e)
 {

    System.err.println("Main Exceptions");

}
return evdDiag;

}//end of query( ) function
Appendix C – Example of REST End Point Path Definition

package web;
import java.util.ArrayList;

// Import java REST libraries
import javax.ws.rs.GET;
import javax.ws.rs.Path;
import javax.ws.rs.PathParam;
import javax.ws.rs.Produces;
import javax.ws.rs.core.MediaType;

// Import Sesame libraries
import org.openrdf.OpenRDFException;

// Import evidence service business objects
import business.QueryRFEBO;
import model.EvidenceDiagnosis;

// Set the REST based query path used to execute the underlying SPARQL query
// The query takes one parameter, a specified RFE id and returns an XML representation of the evidence diagnosis list associated with the RFE
@Path("/query/differentials/{rfe}")
public class Rfe {

    // Specify the return type of the REST query, can be XML or JSON
    // format along with the input parameter taken from the REST url
    // which is the RFE in this case
    @GET
@Produces({MediaType.APPLICATION_XML,
    MediaType.APPLICATION_JSON})

public EvidenceDiagnosis getXML(@PathParam("rfe") String rfe)
throws OpenRDFException {

    // Create evidence diagnosis list object
    EvidenceDiagnosis ed = new EvidenceDiagnosis();

    // Execute the underlying SPARQL command based on the passed
    // RFE returning the XML representation for display at the
    // REST endpoint defined above
    ed = QueryRFEBO.getXML(rfe);
    returned;
    }
}
## Appendix D – Clinical Evidence

### Service REST Endpoints

<table>
<thead>
<tr>
<th>REST URL Path</th>
<th>Data Obtained</th>
</tr>
</thead>
<tbody>
<tr>
<td>interfaces/query/rfes</td>
<td>Return a list of all reasons for encounter defined in the ontology</td>
</tr>
<tr>
<td>interfaces/query/differentials</td>
<td>Return a list of all differential diagnoses defined in the ontology</td>
</tr>
<tr>
<td>interfaces/query/cues</td>
<td>Return a list of all diagnostic cues defined in the ontology</td>
</tr>
<tr>
<td>interfaces/query/signs</td>
<td>Return a list of all diagnostic signs defined in the ontology</td>
</tr>
<tr>
<td>interfaces/query/symptoms</td>
<td>Return a list of all diagnostic symptoms defined in the ontology</td>
</tr>
<tr>
<td>interfaces/query/tests</td>
<td>Return a list of all diagnostic tests defined in the ontology</td>
</tr>
<tr>
<td>interfaces/query/risks</td>
<td>Return a list of all diagnostic risk factors defined in the ontology</td>
</tr>
<tr>
<td>interfaces/query/rfc/quantifications/{rfe}</td>
<td>Return a list of all quantifications which are linked to the specific RFE with ontology identifier {rfe} e.g. AbdominalPainLocalisedOtherRFE</td>
</tr>
<tr>
<td>interfaces/query/rfc/differentials/{rfe}</td>
<td>Return a list of all differential diagnoses which are linked to the specific RFE with ontology identifier {rfe} e.g. AbdominalPainLocalisedOtherRFE</td>
</tr>
<tr>
<td>interfaces/query/rfc/cues /{rfe}</td>
<td>Return a list of all diagnostic cues which are linked to the specific RFE with ontology identifier {rfe} e.g. AbdominalPainLocalisedOtherRFE</td>
</tr>
<tr>
<td>interfaces/query/rfc/signs /{rfe}</td>
<td>Return a list of all diagnostic signs which are linked to the specific RFE with ontology identifier {rfe} e.g. AbdominalPainLocalisedOtherRFE</td>
</tr>
<tr>
<td><strong>interfaces/query/rfes/symptoms /{rfe}</strong></td>
<td>Return a list of all diagnostic symptoms which are linked to the specific RFE with ontology identifier {rfe} e.g. AbdominalPainLocalisedOtherRFE</td>
</tr>
<tr>
<td><strong>interfaces/query/rfes/tests /{rfe}</strong></td>
<td>Return a list of all diagnostic tests which are linked to the specific RFE with ontology identifier {rfe} e.g. AbdominalPainLocalisedOtherRFE</td>
</tr>
<tr>
<td><strong>interfaces/query/rfes/risks /{rfe}</strong></td>
<td>Return a list of all diagnostic risk factors which are linked to the specific RFE with ontology identifier {rfe} e.g. AbdominalPainLocalisedOtherRFE</td>
</tr>
<tr>
<td><strong>interfaces/query/differentials/rfes/{diff}</strong></td>
<td>Return a list of all rfes which are linked to the differential diagnosis with ontology identifier {diff} e.g. UrinaryTractInfection</td>
</tr>
<tr>
<td><strong>interfaces/query/differentials/cues /{diff}</strong></td>
<td>Return a list of all diagnostic cues which are linked to the differential diagnosis with ontology identifier {diff} e.g. UrinaryTractInfection</td>
</tr>
<tr>
<td><strong>interfaces/query/differentials/signs /{diff}</strong></td>
<td>Return a list of all diagnostic signs which are linked to the differential diagnosis with ontology identifier {diff} e.g. UrinaryTractInfection</td>
</tr>
<tr>
<td><strong>interfaces/query/differentials/symptoms /{diff}</strong></td>
<td>Return a list of all diagnostic symptoms which are linked to the differential diagnosis with ontology identifier {diff} e.g. UrinaryTractInfection</td>
</tr>
<tr>
<td><strong>interfaces/query/differentials/tests /{diff}</strong></td>
<td>Return a list of all diagnostic tests which are linked to the differential diagnosis with ontology identifier {diff} e.g. UrinaryTractInfection</td>
</tr>
<tr>
<td>Path</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>interfaces/query/differentials/risks/{diff}</td>
<td>Return a list of all diagnostic risk factors which are linked to the differential diagnosis with ontology identifier {diff} e.g. UrinaryTractInfection</td>
</tr>
<tr>
<td>interfaces/query/differentials/redflags/{diff}</td>
<td>Return a list of all diagnostic red flag cues which are linked to the differential diagnosis with ontology identifier {diff} e.g. OvarianCancer</td>
</tr>
<tr>
<td>interfaces/query/cues/rfes/{cue}</td>
<td>Return a list of all rfes which are linked to the diagnostic cue with ontology identifier {cue} e.g. Fever</td>
</tr>
<tr>
<td>interfaces/query/cues/differentials/{cue}</td>
<td>Return a list of all differential diagnoses which are linked to the diagnostic cue with ontology identifier {cue} e.g. Fever</td>
</tr>
</tbody>
</table>
Appendix E - Definitions of Association Rules Quality Measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
<th>Formulas</th>
</tr>
</thead>
<tbody>
<tr>
<td>A priori or prior probability (AP)</td>
<td>Probability of having positive the condition expressed in the consequent</td>
<td>( \frac{N_+}{N} )</td>
</tr>
<tr>
<td>Variables characterization, to be computed per each variable in the antecedent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posterior probability (PP(\pi))</td>
<td>Probability of having a positive consequent given the variable value is positive (negative). A variable, ( v ), is good discriminator if PP - AP is large</td>
<td>( \frac{N_+ + v \pi}{N_0 \pi} )</td>
</tr>
<tr>
<td>Likelihood ratio (LR(\pi))</td>
<td>Ratio of the probability of variable with result ( x ) in people with positive consequent, and by the probability of antecedent with result ( x ) in people with negative consequent. ( LR = 1 ) is useless; an variable is better discriminatory the more it differs from ( 1 ), that is, greater than ( 1 ) for ( LR_+ ) and lower than ( 1 ) for ( LR_- )</td>
<td>( LR_+ \frac{N_+ + v \pi}{N_+ - v \pi} ) ( LR_- = \frac{1-Sen}{1-Spe} )</td>
</tr>
<tr>
<td>Support (sup)</td>
<td>Proportion of transactions which contain the itemset ( A \cup C ). (Percentage of patients with all previous features)</td>
<td>( \frac{N_+ + a \pi}{N} )</td>
</tr>
<tr>
<td>Lift (or interest) (Lift)</td>
<td>How many times more often ( A ) and ( C ) occur together than expected if they were statistically independent (lift more times appears complain=AbdominalPain, ..., PrimaryCare = ClinicalEnvironment associated to Diagnosis=UrinaryTractInfection, that we could expect if they were independent events.)</td>
<td>( \frac{\text{conf}(A \rightarrow C) \cdot \text{supp}(C)}{\text{conf}(C \rightarrow A) / \text{supp}(a)} = \frac{N_+ + a \pi}{N_+ - a \pi} ) ( \frac{N_+ + v \pi}{N_+ - v \pi} )</td>
</tr>
<tr>
<td>Confidence(or strength) (Conf)</td>
<td>Probability of seeing the rule’s consequent ( C ) under the condition that the transactions also contain the antecedent ( A ) (( \text{conf} ) percentage of the patients having complain=AbdominalPain, ..., PrimaryCare = ClinicalEnvironment, has been diagnosticated with UrinaryTractInfection).</td>
<td>( \frac{\text{supp}(A \rightarrow C) / \text{supp}(C)}{\text{supp}(A) / \text{supp}(a)} = \frac{N_+ + a \pi}{N_+ - a \pi} )</td>
</tr>
<tr>
<td>Conviction (Conv)</td>
<td>Compares the probability that ( A ) appears without ( C ) if they were dependent with the actual frequency of the appearance of ( A ) with ( C ). In that respect it is similar to lift, however, it contrast to lift it is a directional (( \text{conf}(A \rightarrow C) \neq \text{conf}(C \rightarrow A) )) measure since it also uses the information of the absence of the consequent</td>
<td>( (1 - \text{supp}(C)) \cdot (1 - \text{conf}(A \rightarrow C)) = \frac{(1-N_+ + a \pi)}{N_+ + a \pi} )</td>
</tr>
<tr>
<td>Sensitivity (Sen)</td>
<td>Probability of the antecedent being positive given the consequent is positive</td>
<td>( \frac{N_+ + a \pi}{N_+} )</td>
</tr>
<tr>
<td>Specificity (Spe)</td>
<td>Probability of the antecedent being negative given the consequent is negative</td>
<td>( \frac{N_+ - a \pi}{N_-} )</td>
</tr>
</tbody>
</table>
A = association rule antecedent

C= association rule consequent

Rule structure is A -> C

Nc + = number of occurrences of patients with a selected condition as expressed in the rule consequent C

Nc - = number of occurrences of patients without a selected condition as expressed in the rule consequent C

N = total number of patients

V= a selected diagnostic cue or variable V expressed in the rule antecedent
Appendix F – XML Schema for Association Rules

```xml
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="Rule">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="RuleAntecedent" maxOccurs="unbounded"/>
        <xs:element ref="RuleConsecutive"/>
        <xs:element ref="RuleRemark" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="RuleScore" maxOccurs="unbounded"/>
      </xs:sequence>
      <xs:attribute name="RuleScenario" type="xs:string" use="required"/>
      <xs:attribute name="RuleProvenance" type="xs:string" use="required"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="RuleSet">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="Rule" maxOccurs="unbounded"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```
<xs:sequence>
  <xs:attribute name="RuleSetID" type="xs:string" use="required"/>
  <xs:attribute name="RuleSetDate" type="xs:string" use="required"/>
  <xs:attribute name="RuleSetSource" type="xs:string" use="required"/>
  <xs:attribute name="RuleSetComment" type="xs:string" use="required"/>
</xs:complexType>
</xs:element>

<xs:element name="RuleScore">
  <xs:complexType>
    <xs:simpleContent>
      <xs:extension base="xs:string">
        <xs:attribute name="ScoreType" type="xs:string" use="required"/>
      </xs:extension>
    </xs:simpleContent>
  </xs:complexType>
</xs:element>

<xs:element name="RuleRemark" type="xs:string" />
<xs:element name="RuleAntecedent" type="xs:string" />
<xs:complexType>
  <xs:simpleContent>
  </xs:complexType>
</xs:element>
<xs:extensionbase="xs:string">
  <xs:attributename="Description" type="xs:string"
  use="required"/>
  <xs:attributename="AntecendentType" type="xs:string"
  use="required"/>
</xs:extension>
</xs:simpleContent>
</xs:complexType>
</xs:element>
<xs:elementname="RuleConsecutive">
  <xs:complexType>
    <xs:simpleContent>
      <xs:extensionbase="xs:string">
        <xs:attributename="Description" type="xs:string"
        use="required"/>
        <xs:attributename="ConsecutiveEncoding" type="xs:string"
        use="required"/>
      </xs:extension>
    </xs:simpleContent>
  </xs:complexType>
</xs:element>
</xs:schema>
## Appendix G – C# Evidence Client

### Core Classes for Implementing Decision Support Functions

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ClinicalEvidenceOntologyObject</strong></td>
<td>An abstract class which is at the root of all objects extracted from the clinical evidence REST service. At present, it simply handles serialization of objects back to XML.</td>
</tr>
<tr>
<td><strong>ClinicalEvidenceWebInterface</strong></td>
<td>The interface to the TRANSFoRm clinical evidence web service. Defines a number of static methods which may be called to execute queries across the ontology that is stored by the REST service.</td>
</tr>
<tr>
<td><strong>CodeBinding</strong></td>
<td>A code which can be used to specifically identify a particular RFE, cue or diagnosis.</td>
</tr>
<tr>
<td><strong>Cue</strong></td>
<td>A cue is data that is recorded by the clinician in session with the patient. It could be a symptom or perhaps some observation of the patient’s lifestyle. Ultimately, a cue will be used to identify a probable diagnosis. In future implementations, cue will have a number of child classes which refer to specific types of cues such as symptom, risk, sign, etc. For now, these children are dissociated classes.</td>
</tr>
<tr>
<td><strong>CueDifferentialsList</strong></td>
<td>An indexable collection of evidence diagnosis. Required for serialisation, but will be replaced by EvidenceDiagnosisList or something more general at a later date.</td>
</tr>
<tr>
<td><strong>CueList</strong></td>
<td>A serializable list of EvidenceDiagnosis objects. Has an indexer as a transparent layer.</td>
</tr>
<tr>
<td>Class/Concept</td>
<td>Description</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Demographic</td>
<td>A demographic describes some contextual aspect of a patient's condition, such as age, weight, country, etc.</td>
</tr>
<tr>
<td>DifferentialCueMap</td>
<td>Maintains a mapping between differential identifiers and lists of cues associated with a particular differential</td>
</tr>
<tr>
<td>DifferentialCueMap.DiffCueItem</td>
<td>A temporary class. This is required to serialize and deserialize a dictionary to and from XML as C# does not support this operation natively. This class represents a single key value pair in the dictionary.</td>
</tr>
<tr>
<td>EvidenceDiagnosis</td>
<td>An evidence diagnosis is a specific diagnosis of a patient's condition based on an examination of the various cues and rfeas that are present</td>
</tr>
<tr>
<td>EvidenceDiagnosisList</td>
<td>A serializable list of EvidenceDiagnosis objects. Has an indexer as a transparent layer between the calling function and the underlying list container.</td>
</tr>
<tr>
<td>PatientCase</td>
<td>A class which handles a single patient's case. Aggregates information about the patient's symptoms and uses it to generate lists of potential diagnoses</td>
</tr>
<tr>
<td>PatientEvidenceSet</td>
<td>A collection of evidence gathered about a patient's condition</td>
</tr>
<tr>
<td>Quantification</td>
<td>A quantification is a measure of how strongly related a RFE is to a particular differential diagnosis</td>
</tr>
<tr>
<td>QuantificationContext</td>
<td>The context of the patient described by the quantification, e.g. sex, age, country of origin</td>
</tr>
<tr>
<td>QuantificationList</td>
<td>A serializable list of Quantification objects. Has</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>QuantificationValue</td>
<td>A specific value that is associated with a quantification.</td>
</tr>
<tr>
<td>Rfe</td>
<td>A reason for encounter (RFE) is the reason for an initial interaction between the patient and the clinician. It is the reason why the patient presented themselves to the clinic for examination.</td>
</tr>
<tr>
<td>RfeList</td>
<td>A serializable list of Rfe objects. Has an indexer as a transparent layer between the calling function and the underlying list container.</td>
</tr>
<tr>
<td>Risk</td>
<td>A risk describes some condition of the patient's lifestyle or history which may make them more susceptible to a particular disease, e.g. family history of heart disease, long term smoker, etc.</td>
</tr>
<tr>
<td>RisksList</td>
<td>A serializable list of Risk objects. Has an indexer as a transparent layer between the calling function and the underlying list container.</td>
</tr>
<tr>
<td>Sign</td>
<td>A sign is a cue which is measured by the clinician, e.g. temperature.</td>
</tr>
<tr>
<td>SignsList</td>
<td>A serializable list of Sign objects. Has an indexer as a transparent layer between the calling function and the underlying list container.</td>
</tr>
<tr>
<td>Symptom</td>
<td>A symptom is a cue which the clinician observes, e.g. pale skin, cough, etc. A symptom may be the same as a reason for encounter, however, in the knowledge ontology, the two are not linked.</td>
</tr>
<tr>
<td><strong>SymptomList</strong></td>
<td>A serializable list of Symptom objects. Has an indexer as a transparent layer between the calling function and the underlying list container.</td>
</tr>
<tr>
<td><strong>Test</strong></td>
<td>A test is a type of cue which describes how to check for a particular diagnosis. For example, it may suggest a dipstick urinalysis test in order to check for the presence of the disease.</td>
</tr>
<tr>
<td><strong>TestList</strong></td>
<td>A serializable list of Test objects. Has an indexer as a transparent layer between the calling function and the underlying list container.</td>
</tr>
<tr>
<td><strong>Demographic</strong></td>
<td>A demographic describes some contextual aspect of a patient's condition, such as age, weight, country, etc.</td>
</tr>
<tr>
<td><strong>DifferentialCueMap</strong></td>
<td>Maintains a mapping between differential identifiers and lists of cues associated with a particular differential.</td>
</tr>
<tr>
<td><strong>DifferentialCueMap.DiffCueItem</strong></td>
<td>A temporary class. This is required to serialize and deserialize a dictionary to and from XML as C# does not support this operation natively. This class represents a single key value pair in the dictionary.</td>
</tr>
<tr>
<td><strong>EvidenceDiagnosis</strong></td>
<td>An evidence diagnosis is a specific diagnosis of a patient's condition based on an examination of the various cues and rfes that are present.</td>
</tr>
<tr>
<td><strong>EvidenceDiagnosisList</strong></td>
<td>A serializable list of EvidenceDiagnosis objects. Has an indexer as a transparent layer between the calling function and the underlying list container.</td>
</tr>
<tr>
<td><strong>PatientCase</strong></td>
<td>A class which handles a single patient's case. Aggregates information about the patients.</td>
</tr>
</tbody>
</table>
symptoms and uses them to generate lists of potential diagnoses
## Appendix H – C# Evidence Client

Methods Implemented by the ClinicalEvidenceWebInterface class

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>getAssociatedDifferentialCues</td>
<td>Gets the differentials that are associated with a particular RFE. Then retrieves the cues that are associated with each differential. The result is returned as a dictionary which is indexed by the name of a differential, and the values are lists of cues that are associated with that particular differential.</td>
</tr>
<tr>
<td>GetCueDifferentials</td>
<td>Get a list of differential diagnosis that are associated with a particular cue. Returns an indexable list of differential diagnosis.</td>
</tr>
<tr>
<td>GetCueRfes</td>
<td>Get a list of all reasons for encounter associated with a particular cue.</td>
</tr>
<tr>
<td>GetDifferentialCues</td>
<td>Get a list of cues that are associated with a particular differential diagnosis. Results are returned as an indexable object.</td>
</tr>
<tr>
<td>GetDifferentialRedFlags</td>
<td>Get a list of symptoms which are associated with a differential diagnosis that are considered strong indicators of its presence. Returns an indexable list of symptoms.</td>
</tr>
<tr>
<td>GetDifferentialRfes</td>
<td>Get a list of reasons for encounter that are associated with a particular differential diagnosis.</td>
</tr>
<tr>
<td>GetDifferentialRisks</td>
<td>Get a list of risks that are associated with a particular evidence diagnosis. The results are returned as an indexable object.</td>
</tr>
<tr>
<td>Function</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>GetDifferentialSigns</td>
<td>Gets a list of signs that are associated with a particular evidence diagnosis. The results are returned as an indexable object.</td>
</tr>
<tr>
<td>GetDifferentialSymptoms</td>
<td>Get a list of the symptoms that are associated with a particular differential diagnosis. The results are returned as an indexable list of objects.</td>
</tr>
<tr>
<td>GetDifferentialTests</td>
<td>Get a list of tests that can be performed to check for a particular differential diagnosis. The results are returned as an indexable list of objects.</td>
</tr>
<tr>
<td>GetRfeCues</td>
<td>Get a list of all cues associated with a reason for encounter.</td>
</tr>
<tr>
<td>GetRfeDifferentials</td>
<td>Get the differentials (possible diagnosis) for a reason for encounter.</td>
</tr>
<tr>
<td>GetRfeQuantifications</td>
<td>Get a list of the quantifications that are associated with the RFE that is passed as an argument. Result is returned as an indexable object.</td>
</tr>
<tr>
<td>GetRfeRisks</td>
<td>Get a list of all risks associated with a reason for encounter.</td>
</tr>
<tr>
<td>GetRfeSigns</td>
<td>Get a list of all signs associated with a reason for encounter.</td>
</tr>
<tr>
<td>GetRfeSymptoms</td>
<td>Get a list of all symptoms associated with a reason for encounter.</td>
</tr>
<tr>
<td>GetRfeTests</td>
<td>Get a list of all tests associated with a reason for encounter.</td>
</tr>
<tr>
<td>GetRiskDifferentials</td>
<td>Get a list of differential diagnosis associated with a particular risk.</td>
</tr>
<tr>
<td>GetRiskRfes</td>
<td>Get a list of all reasons for encounter associated with a particular risk.</td>
</tr>
<tr>
<td>GetSignDifferentials</td>
<td>Get a list of differential diagnosis associated with a particular sign.</td>
</tr>
<tr>
<td>GetSignRfes</td>
<td>Get a list of all reasons for encounter associated with a particular sign.</td>
</tr>
<tr>
<td>GetSymptomDifferentials</td>
<td>Get a list of differential diagnosis associated with a particular symptom diagnosis.</td>
</tr>
<tr>
<td>Function</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>GetSymptomRfes</td>
<td>Get a list of all reasons for encounter associated with a particular symptom</td>
</tr>
<tr>
<td>GetTestDifferentials</td>
<td>Get a list of differential diagnosis associated with a particular test</td>
</tr>
<tr>
<td>GetTestRfes</td>
<td>Get a list of all reasons for encounter associated with a particular test</td>
</tr>
<tr>
<td>ListCues</td>
<td>Get a list of all cues in the clinical evidence ontology</td>
</tr>
<tr>
<td>ListDifferentials</td>
<td>Get a list of all differential diagnosis in the clinical evidence ontology</td>
</tr>
<tr>
<td>ListRfes</td>
<td>Retrieves a list of all RFEs that are stored in the clinical evidence ontology</td>
</tr>
<tr>
<td>ListRisks</td>
<td>Get a list of all risks in the clinical evidence ontology</td>
</tr>
<tr>
<td>ListSigns</td>
<td>Get a list of all signs in the clinical evidence ontology</td>
</tr>
</tbody>
</table>
## Appendix I – C# Evidence Client

### Methods Implemented by the PatientCase Class

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UpdateCase(PatientEvidenceSet)</strong></td>
<td>Accept patient evidence set data from the CDSS relating to a patient case as a new evidence set. Use this data to re-filter and re-rank the returned list of possible diagnoses to consider to the CDSS interface.</td>
</tr>
</tbody>
</table>
Appendix J – The PatientEvidenceSet C# Class

```csharp
public partial class PatientEvidenceSet {
    private List<RFE> rFESField;
    private List<Demographic> demographicsField;
    private List<Risk> risksField;
    private List<Sign> signsField;
    private List<Symptom> symptomsField;
    private ushort patientIDField;

    public PatientEvidenceSet() {
        this.symptomsField = new List<Symptom>();
        this.signsField = new List<Sign>();
        this.risksField = new List<Risk>();
        this.demographicsField = new List<Demographic>();
        this.rFESField = new List<RFE>();
    }

    public List<RFE> RFES {
        get {
            return this.rFESField;
        }
        set {
            this.rFESField = value;
        }
    }

    public List<Demographic> Demographics {
        get {
            return this.demographicsField;
        }
        set {
            this.demographicsField = value;
        }
    }

    public List<Risk> Risks {
    }
```

Appendix K – Example Output of Systematic Review of Literature to Support Diagnosis of Urinary Tract Infection

<table>
<thead>
<tr>
<th>Reference</th>
<th>Type of study</th>
<th>Population</th>
<th>Location</th>
<th>Exclusions</th>
<th>Demographic</th>
<th>Quality</th>
<th>Empirical Results</th>
<th>Related Study (Review)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bent et al 2002¹</td>
<td>Meta Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference</th>
<th>Type of study</th>
<th>Population</th>
<th>Location</th>
<th>Exclusions</th>
<th>Demographic</th>
<th>Quality</th>
<th>Empirical Results</th>
<th>Related Study (Review)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bent et al 2002¹</td>
<td>Meta Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Rated Level 1 (best) to Level 5 (poor) based on independent review by authors using JAMA evidence methodology. See * below.

JAMA Evidence Summary of key UTI papers to 2001 – Pooled and Combined Likelihood Ratios
<table>
<thead>
<tr>
<th>Reference</th>
<th>Type of study</th>
<th>Population</th>
<th>Location</th>
<th>Exclusions</th>
<th>Demographic</th>
<th>Quality</th>
<th>Empirical Results</th>
<th>Related Study (Review)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gallagher et al 1965&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Primary</td>
<td>130 patients over 8 months with symptoms of UTI</td>
<td>8 GP practices in New Zealand</td>
<td>92% of UTI confirmed women (71 of 77 cases) Ages 1-80</td>
<td>JAMA Level 1</td>
<td>Presenting symptoms by % frequency for confirmed UTI and negative UTI patients</td>
<td>Bent et al 2002&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Lawson et al 1973&lt;sup&gt;3&lt;/sup&gt;</td>
<td>Primary</td>
<td>343 females over 18 months with symptoms of UTI</td>
<td>2 GP practices in UK Pregnant or diabetic patients</td>
<td>Ages 15-55</td>
<td>JAMA Level 1</td>
<td>Presenting symptoms by % frequency for confirmed UTI patients Incidence rates in two GP practices</td>
<td>Bent et al 2002&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Type of study</td>
<td>Population</td>
<td>Location</td>
<td>Exclusions</td>
<td>Demographic</td>
<td>Quality</td>
<td>Empirical Results</td>
<td>Related Study (Review)</td>
</tr>
<tr>
<td>------------</td>
<td>---------------</td>
<td>------------</td>
<td>----------</td>
<td>------------</td>
<td>-------------</td>
<td>---------</td>
<td>-------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Nazareth et al 1993⁴</td>
<td>Primary</td>
<td>Women presenting over 3 months with symptoms of UTI</td>
<td>2 GP practices in London, UK</td>
<td>Antibiotics prescribed in previous 4 weeks to presentation, Pregnant women or any concurrent medical condition</td>
<td>Women ages 16-45</td>
<td>JAMA Level 1</td>
<td>Presenting symptoms by % frequency for confirmed UTI and negative UTI patients, Demographic data for confirmed UTI and negative UTI patients</td>
<td>Bent et al 2002¹</td>
</tr>
<tr>
<td>Reference</td>
<td>Type of study</td>
<td>Population</td>
<td>Location</td>
<td>Exclusions</td>
<td>Demographic</td>
<td>Quality</td>
<td>Empirical Results</td>
<td>Related Study (Review)</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------</td>
<td>------------</td>
<td>----------</td>
<td>----------------------------------------------------------------------------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>Gupta et al 2001⁵</td>
<td>Primary</td>
<td>172 Women 18 or older with at least 2 UTI's in previous 12 months</td>
<td>US University-based primary health care clinic.</td>
<td>Women with a known allergy to fluoroquinolones; a full-term pregnancy in the past 12 months; or a history of diabetes, hypertension, or renal disease</td>
<td>Women ages 18-51 (Mean 23)</td>
<td>JAMA Level 5</td>
<td>Diagnostic accuracy and outcomes of 172 patient-initiated treatment episodes of recurrent urinary tract infection</td>
<td>Bent et al 2002¹</td>
</tr>
<tr>
<td>Wong et al 1985⁶</td>
<td>Primary</td>
<td>53 women presenting with UTI symptoms or urinary and vaginal complaints</td>
<td>US based STD clinic</td>
<td>Women with vaginitis or cervicitis</td>
<td>Not given</td>
<td>JAMA Level 4</td>
<td>Presenting symptoms by % frequency for confirmed UTI patients.</td>
<td>Bent et al 2002¹</td>
</tr>
<tr>
<td>Reference</td>
<td>Type of study</td>
<td>Population</td>
<td>Location</td>
<td>Exclusions</td>
<td>Demographic</td>
<td>Quality</td>
<td>Empirical Results</td>
<td>Related Study (Review)</td>
</tr>
<tr>
<td>----------------------------</td>
<td>---------------</td>
<td>------------</td>
<td>----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------</td>
<td>---------</td>
<td>-------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Medina Bombardo et al 2003</td>
<td>Primary</td>
<td>417 Women</td>
<td>18 Balearic Islands GP</td>
<td>Women treated for UTI in previous 15 days prior to presentation</td>
<td>Women 15-90, average age 44</td>
<td></td>
<td></td>
<td>Presenting symptoms analysed to provide frequency, sensitivity, specificity, unadjusted LR+, LR- with 95% CI. Symptoms by previous medical history analysed to provide frequency, sensitivity, specificity, unadjusted LR+, LR- with 95% CI.</td>
</tr>
<tr>
<td>Reference</td>
<td>Type of study</td>
<td>Population</td>
<td>Location</td>
<td>Exclusions</td>
<td>Demographic</td>
<td>Quality</td>
<td>Empirical Results</td>
<td>Related Study (Review)</td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------</td>
<td>--------------------------------------------</td>
<td>-------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>--------------</td>
<td>---------</td>
<td>-------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Scholes et al 2000⁸</td>
<td>Primary</td>
<td>229 Women with Recurrent UTIs (3 UTIs in last year or 2 UTIs in last 6 months)</td>
<td>Seattle based university health centre and health maintenance organisation (HMO)</td>
<td>episodes of asymptomatic bacteriuria and symptomatic but non culture confirmed UTI cases</td>
<td>Women 18-30</td>
<td></td>
<td>Multivariate risk factors associated with recurrent UTI with odds ratios and 95% CI</td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Type of study</td>
<td>Population</td>
<td>Location</td>
<td>Exclusions</td>
<td>Demographic</td>
<td>Quality</td>
<td>Empirical Results</td>
<td>Related Study (Review)</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------</td>
<td>------------</td>
<td>----------</td>
<td>------------</td>
<td>-------------</td>
<td>---------</td>
<td>-------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Schmieman n et al 2010</td>
<td>Review</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No evidence level graded against articles</td>
<td>Summary of other studies indicating diagnostic accuracy of dipstick test with sensitivity, specificity, LR+, LR-</td>
<td>Summary of diagnostic UTI algorithms from other studies with sensitivity and specificity</td>
</tr>
<tr>
<td>Reference</td>
<td>Type of study</td>
<td>Population</td>
<td>Location</td>
<td>Exclusions</td>
<td>Demographic</td>
<td>Quality</td>
<td>Empirical Results</td>
<td>Related Study (Review)</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------</td>
<td>-------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>------------------------------</td>
<td>---------</td>
<td>------------------------------------------------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Heckerling et al 2007(^\text{10})</td>
<td>Primary</td>
<td>212 Women presenting with symptoms of UTI between February 1997 and September 1999</td>
<td>Ambulatory university care clinic in Omaha</td>
<td>Women complaining of vaginal discharge or bleeding women with a known anatomic or functional abnormality of the urinary tract, such as a neurogenic bladder, an indwelling urinary catheter, or recent urinary tract instrumentation</td>
<td>Women aged 19 – 84, mean age 41</td>
<td></td>
<td>Presenting symptoms analysed to provide, unadjusted +LR with 95% CI.</td>
<td>Schmiemann et al 2010</td>
</tr>
<tr>
<td>Reference</td>
<td>Type of study</td>
<td>Population</td>
<td>Location</td>
<td>Exclusions</td>
<td>Demographic</td>
<td>Quality</td>
<td>Empirical Results</td>
<td>Related Study (Review)</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------</td>
<td>------------</td>
<td>----------</td>
<td>----------------------------------------------------------------------------</td>
<td>----------------------</td>
<td>---------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>McIsaac et al 2002&lt;sup&gt;11&lt;/sup&gt;</td>
<td>Primary</td>
<td>231 women presenting with symptoms of UTI between January 1, 1998, and January 7, 2000.</td>
<td>4 urban academic family medicine Clinics in Toronto, Canada</td>
<td>taking antibiotics in the previous 7 days, pregnant, were following up a previously diagnosed UTI.</td>
<td>Women aged 16 or older</td>
<td></td>
<td>Presenting symptoms and risk factors by % frequency for confirmed UTI CPR for deciding on antibiotic treatment</td>
<td>Schmiemann et al 2010</td>
</tr>
<tr>
<td>Reference</td>
<td>Type of study</td>
<td>Population</td>
<td>Location</td>
<td>Exclusions</td>
<td>Demographic</td>
<td>Quality</td>
<td>Empirical Results</td>
<td>Related Study (Review)</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------</td>
<td>----------------------</td>
<td>-------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------------------</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>McIsaac et al 2007</td>
<td>Primary</td>
<td>331 females with suspected cystitis</td>
<td>225 Canadian family physicians</td>
<td>children younger than 16 years, pregnant women, nursing home residents, immunocompromised patients, taking antibiotics or with renal tract abnormalities, indwelling catheters, inability to understand English.</td>
<td>Women aged 16-99, mean age 45.2 years</td>
<td></td>
<td>Presenting signs and symptoms by % frequency for confirmed UTI</td>
<td>Schmiemann et al 2010</td>
</tr>
</tbody>
</table>

Adjusted 3 variable CPR (adjusted from original McIsaac et al 2002)
<table>
<thead>
<tr>
<th>Reference</th>
<th>Type of study</th>
<th>Population</th>
<th>Location</th>
<th>Exclusions</th>
<th>Demographic</th>
<th>Quality</th>
<th>Empirical Results</th>
<th>Related Study (Review)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Little et al 2006¹³</td>
<td>Primary</td>
<td>427 women presenting with symptoms of UTI</td>
<td>117 GPs and practice nurses from 67 practices in the south of England</td>
<td>Patients where other diagnoses were considered to be likely, for example, pregnant people, people aged over 70, People with current severe mental problems (such as dementia)</td>
<td>Women aged 18-70</td>
<td>CPR predicting diagnosis of UTI based on selected signs/symptoms/tests</td>
<td>Schmiemann et al 2010</td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Type of study</td>
<td>Population</td>
<td>Location</td>
<td>Exclusions</td>
<td>Demographic</td>
<td>Quality</td>
<td>Empirical Results</td>
<td>Related Study (Review)</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------</td>
<td>-------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------</td>
<td>---------</td>
<td>-------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Little et al 2010&lt;sup&gt;14&lt;/sup&gt;</td>
<td>Primary</td>
<td>434 adult females with suspected lower UTI</td>
<td>117 primary care clinicians (doctors or practice nurses) from 62 practices in the south of England</td>
<td>Patients where other diagnoses were considered to be likely, for example, women with vaginal symptoms, pregnant women, people aged over 70, People with current severe mental problems (such as dementia)</td>
<td>Women aged 18-70 years</td>
<td></td>
<td>Modified CPR based on Little et al&lt;sup&gt;13&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Type of study</td>
<td>Population</td>
<td>Location</td>
<td>Exclusions</td>
<td>Demographic</td>
<td>Quality</td>
<td>Empirical Results</td>
<td>Related Study (Review)</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------</td>
<td>------------</td>
<td>----------</td>
<td>------------</td>
<td>-------------</td>
<td>---------</td>
<td>---------------------------------------------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Fahey et al 2003&lt;sup&gt;15&lt;/sup&gt;</td>
<td>Primary</td>
<td>135 women presenting with symptoms of UTI</td>
<td>29 GPs in 8 Bristol, UK practices</td>
<td>Pregnant women, Men</td>
<td></td>
<td></td>
<td>Relationship of 11 symptoms to probability of receiving testing or antibiotic treatment</td>
<td>LR+ with 95% CI for 11 symptoms against two ‘gold standards’ – positive urine culture and returning with symptoms within 1 month</td>
</tr>
<tr>
<td>Reference</td>
<td>Type of study</td>
<td>Population</td>
<td>Location</td>
<td>Exclusions</td>
<td>Demographic</td>
<td>Quality</td>
<td>Empirical Results</td>
<td>Related Study (Review)</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------</td>
<td>------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>---------</td>
<td>-------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Hooton et al 1996¹⁶</td>
<td>Primary</td>
<td>796 healthy women</td>
<td>University of Washington Student Health Center and the Group Health Cooperative of Puget Sound</td>
<td>More than 1 UTI in previous 12 months, Pregnant women or planned pregnancy in 6 months of study, Chronic illness requiring medical supervision, Used systemic antimicrobial agents within the previous 14 days, Known anatomical abnormality of the urinary tract.</td>
<td>Women aged 18-40 starting new method of contraception (or started in last 6 weeks)</td>
<td></td>
<td>Incidence of UTI by population characteristics</td>
<td>Relative risk of UTI based on chosen selected variables including contraception with 95% CI</td>
</tr>
<tr>
<td>Reference</td>
<td>Type of study</td>
<td>Population</td>
<td>Location</td>
<td>Exclusions</td>
<td>Demographic</td>
<td>Quality</td>
<td>Empirical Results</td>
<td>Related Study (Review)</td>
</tr>
<tr>
<td>----------------------------</td>
<td>--------------</td>
<td>----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>--------------------------------</td>
<td>-----------------------------------</td>
<td>-------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Hummers-Pradier et al 2005</td>
<td>Primary</td>
<td>445 adult women presenting with symptoms of UTI including those with risk factors, comorbidity or recent antibiotic treatment</td>
<td>36 teaching general practices of the Department of General Practice, University of Göttingen</td>
<td>patients with an obvious other diagnosis explaining their symptoms (i.e. vaginitis)</td>
<td>Women aged 33-71 (median age 53)</td>
<td>Presenting signs and symptoms and risk factors by % frequency for suspected UTI</td>
<td>Presenting signs and symptoms and risk factors by % frequency for suspected UTI</td>
<td>GPs management of patients analysed by % for suspected UTI</td>
</tr>
<tr>
<td>Probability</td>
<td>Factors predicting UTI with odds ratios and 95% CI.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>--------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**JAMA Evidence Level Definitions**

Level 1 evidence refers to a primary, prospective study of the accuracy or precision of the clinical examination in the relevant clinical condition. For studies dealing with accuracy, this requires independent, blind comparisons of clinical findings with a criterion standard (or gold standard) of diagnosis or etiology among a large number (>50) of consecutive patients suspected of having the relevant clinical condition. For studies dealing with precision, this requires 2 or more independent blinded raters of symptoms or signs in a large number of patients suspected of having the relevant clinical condition.

Level 2 studies were analogous to level 1 studies but with smaller numbers of patients (10-50), widening the confidence limits of the resulting calculations.

Level 3 studies were based on a retrospective design (i.e., clinical findings determined by chart review).

Level 4 studies included non-consecutive patients, generally selected because of their definitive results for the findings under study, or a non-blinded comparison of clinical findings with a gold standard.

Level 5 studies included studies with an uncertain gold standard or a poorly defined study population.
Appendix L – Full Published Papers from this Research


A Multi-step Maturity Model for the Implementation of Electronic and Computable Diagnostic Clinical Prediction Rules (eCPRs)

Derek Corrigan  
*HRB Centre for Primary Care Research, RCSI Medical School, Dublin, derekcorrigan@rcsi.ie*

Ronan McDonnell  
*HRB Centre for Primary Care Research, RCSI Medical School, Dublin, rmcdonnell@rcsi.ie*

Atieh Zarabzadeh  
*HRB Centre for Primary Care Research, RCSI Medical School, Dublin, atiehzarabzadeh@rcsi.ie*

Tom Fahey  
*HRB Centre for Primary Care Research, RCSI Medical School, Dublin, tomfahey@rcsi.ie*

Follow this and additional works at: [http://repository.academyhealth.org/egems](http://repository.academyhealth.org/egems)

Recommended Citation  
Corrigan, Derek; McDonnell, Ronan; Zarabzadeh, Atieh; and Fahey, Tom (2015) "A Multi-step Maturity Model for the Implementation of Electronic and Computable Diagnostic Clinical Prediction Rules (eCPRs)," *eGEMs (Generating Evidence & Methods to improve patient outcomes)*: Vol. 3: Iss. 2, Article 8.  
DOI: [http://dx.doi.org/10.13063/2327-9214.1153](http://dx.doi.org/10.13063/2327-9214.1153)  
Available at: [http://repository.academyhealth.org/egems/vol3/iss2/8](http://repository.academyhealth.org/egems/vol3/iss2/8)

This Informatics Model/Framework is brought to you for free and open access by the EDM Forum Products and Events at EDM Forum Community. It has been peer-reviewed and accepted for publication in eGEMs (Generating Evidence & Methods to improve patient outcomes).

The Electronic Data Methods (EDM) Forum is supported by the Agency for Healthcare Research and Quality (AHRQ), Grant 1U18HS022789-01. eGEMs publications do not reflect the official views of AHRQ or the United States Department of Health and Human Services.
A Multi-step Maturity Model for the Implementation of Electronic and Computable Diagnostic Clinical Prediction Rules (eCPRs)

Abstract

Introduction: The use of Clinical Prediction Rules (CPRs) has been advocated as one way of implementing actionable evidence-based rules in clinical practice. The current highly manual nature of deriving CPRs makes them difficult to use and maintain. Addressing the known limitations of CPRs requires implementing more flexible and dynamic models of CPR development. We describe the application of Information and Communication Technology (ICT) to provide a platform for the derivation and dissemination of CPRs derived through analysis and continual learning from electronic patient data.

Model Components: We propose a multistep maturity model for constructing electronic and computable CPRs (eCPRs). The model has six levels – from the lowest level of CPR maturity (literature-based CPRs) to a fully electronic and computable service-oriented model of CPRs that are sensitive to specific demographic patient populations. We describe examples of implementations of the core model components – focusing on CPR representation, interoperability, electronic dissemination, CPR learning, and user interface requirements.

Conclusion: The traditional focus on derivation and narrow validation of CPRs has severely limited their wider acceptance. The evolution and maturity model described here outlines a progression toward eCPRs consistent with the vision of a learning health system (LHS) – using central repositories of CPR knowledge, accessible open standards, and generalizable models to avoid repetition of previous work. This is useful for developing more ambitious strategies to address limitations of the traditional CPR development life cycle. The model described here is a starting point for promoting discussion about what a more dynamic CPR development process should look like.

Acknowledgements
This work was partially funded by the Health Research Board (HRB) of Ireland through the HRB Centre for Primary Care Research under Grant HRC/2007/1. TRANSFoRm is partially funded by the European Commission – DG INFSO (FP7 247787).

Keywords
Evidence Based Medicine, Health Information Technology, Research Translation, Clinical Prediction Rules, Learning Health System

Disciplines
Health Information Technology

Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.

This model/framework is available at EDM Forum Community: http://repository.academyhealth.org/egems/vol3/iss2/8
Introduction: The use of Clinical Prediction Rules (CPRs) has been advocated as one way of implementing actionable evidence-based rules in clinical practice. The current highly manual nature of deriving CPRs makes them difficult to use and maintain. Addressing the known limitations of CPRs requires implementing more flexible and dynamic models of CPR development. We describe the application of Information and Communication Technology (ICT) to provide a platform for the derivation and dissemination of CPRs derived through analysis and continual learning from electronic patient data.

Model Components: We propose a multistep maturity model for constructing electronic and computable CPRs (eCPRs). The model has six levels – from the lowest level of CPR maturity (literature-based CPRs) to a fully electronic and computable service-oriented model of CPRs that are sensitive to specific demographic patient populations. We describe examples of implementations of the core model components – focusing on CPR representation, interoperability, electronic dissemination, CPR learning, and user interface requirements.

Conclusion: The traditional focus on derivation and narrow validation of CPRs has severely limited their wider acceptance. The evolution and maturity model described here outlines a progression toward eCPRs consistent with the vision of a learning health system (LHS) – using central repositories of CPR knowledge, accessible open standards, and generalizable models to avoid repetition of previous work. This is useful for developing more ambitious strategies to address limitations of the traditional CPR development life cycle. The model described here is a starting point for promoting discussion about what a more dynamic CPR development process should look like.
Introduction

Clinical Prediction Rules as a Form of Evidence-Based Medicine (EBM)

Evidence-based medicine (EBM) has long been advocated as one way of supporting diagnostic reasoning that is based on a more rigorous and systematic approach.1-2

One form of EBM is the Clinical Prediction Rule (CPR).3-5 CPRs are typically derived through conducting manually intensive observational studies that elicit quantified epidemiological associations using statistical or probabilistic techniques.5 The current highly manual nature of deriving CPRs also makes them difficult to use and maintain.6 With some exceptions the format for CPRs dissemination has traditionally been literature based, putting an onus on clinicians to search literature for suitable CPRs.7

Addressing the known limitations of CPRs requires implementing more flexible and dynamic models of CPR development. We describe the application of Information and Communication Technology (ICT) to provide a platform for derivation and dissemination of CPRs derived through analysis and continual learning from electronic patient data. We present an incremental model of CPR development that is of interest to those in the clinical research community advocating wider use of evidence-based CPRs through translational research, and to those designing or implementing service-oriented, rule-based decision support systems (DSS).

Limitations of Traditional CPR Development

Despite the existence of an accepted development methodology for producing CPRs (Table 1), the development of many CPRs focuses on the derivation phase (Level 1) of the CPR life cycle,8 lacking subsequent validation (Levels 2 and 3) and impact analysis (Level 4).9

This lack of validation for many CPRs limits their applicability to the same patient populations used for the original derivation. Scores may vary when the CPR is applied to populations with gender, age, or clinical settings that are different from the original derivation population. The Alvarado score, for example, has been found to perform best in adult male populations.10 This has implications for the applicability of any published CPR as changes take

Table 1. Accepted CPR Development Methodology

<table>
<thead>
<tr>
<th>CPR CATEGORY</th>
<th>LEVEL OF EVIDENCE REQUIRED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derivation (Level 1)</td>
<td>Factors with predictive power are identified in order to base the rule on a derivation patient population.</td>
</tr>
<tr>
<td>Narrow validation (Level 2)</td>
<td>The rule is applied to a different patient set with characteristics similar to the original derivation population.</td>
</tr>
<tr>
<td>Broad validation (Level 3)</td>
<td>The rule is applied to another population with different characteristics from the original derivation population.</td>
</tr>
<tr>
<td>Impact analysis (Level 4)</td>
<td>The impact of the rule may be tested and assessed in terms of its effect on clinical outcomes, physician behavior, or costs.</td>
</tr>
</tbody>
</table>
place over time in the demographics of the original rule-derivation study population.

**Existing Implementations of CPRs in Decision Support Systems (DSSs)**

There are a number of additional barriers to consider when implementing CPRs electronically as part of DSSs. Previous attempts to deploy CPRs highlight additional issues to consider in implementing clinical DSSs, including the following:

- Validation and impact analysis of CPRs is restricted due to lack of connectivity to wider patient populations beyond the original electronic tools to which they are initially deployed and tied.\(^{11-13}\)
- With some exceptions where evidence is disseminated using open standards, and separate from the application itself,\(^{14-19}\) decision support tools are tied to specific proprietary clinical systems, which lack support for wider Electronic Health Record (EHR) workflow integration across other systems.
- Implementations of decision support tools focus on individual CPR models and are not easily portable to implement other CPRs, resulting in redevelopment efforts for each implemented rule.\(^{11-13,20-26}\)
- Successful implementations and clinical acceptance of the deployed tools necessitate a collaborative multidisciplinary approach to define the nature of the intervention required and the actual workflow of the CPR in practice.\(^{11,21-23}\)

These limitations can be considered part of a wider problem of successfully translating clinical research knowledge into clinical practice using ICT tools.

**CPRs and the Learning Health System (LHS)**

Rapidly translating clinical knowledge into practice is a core objective of the learning health system (LHS).

The current traditional model of CPR development is considered to be at a low level of technological development with respect to what has been termed “the pyramid of evidence.”\(^{27-28}\) Research initiatives have defined what should constitute the core components of the LHS.\(^{29-30}\) Within a virtuous cycle of health improvement a number of important requirements have been identified to support this knowledge translation capability, including the following:

- Generating **valid clinical knowledge**;
- **Packaging and curating knowledge** so it is widely accessible and actionable, and putting knowledge to use to effect change;
- Developing **meaningful use of the EHR** to support diagnostic and therapeutic support based on evidence;
- Developing a computable representation of research evidence and making that **available to EHR systems as a Web service**; and
- Developing a means of providing diagnostic or therapeutic prompts within an EHR that **works across a variety of EHR systems**.\(^{30}\)

We propose the implementation of LHS knowledge-translation capabilities in the constructing of electronic and computable CPRs (eCPRs). The eCPR can be considered an evolution of the traditional CPR development methodology that moves CPR development toward the top of the traditional “pyramid of evidence.” It provides for the electronic derivation and dissemination of CPRs that are computable, updateable, and versionable based on continuous learning obtained from analysis of underlying derivation data. This platform implements model-based and service-oriented architectures, using open interoperability standards to exploit the potential of data mining of aggregated sources of EHRs for CPR development.
Components of the eCPR Maturity Model

The multistep maturity model for eCPR implementation consists of six incremental levels, as shown in Figure 1 and described below. The model can be used to assess the current level of development for an organization using CPRs, and how the organization might develop it further.

Each model level also describes interoperability characteristics that it supports. The definition of interoperability we are using is as provided by the Office of the National Coordinator for Health IT.31 This definition describes four interoperability layers:

- **Syntax**: content and structure;
- **Semantics**: vocabulary and code – sets and terminology;
- **Transport**: method by which information is moved from system to system; and
- **Services**: the infrastructure components deployed and used to accomplish specific information exchange objectives.

**Level 1: Literature-Based CPRs**

Interoperability Layers: *Not interoperable – stand-alone tool.*

A literature search of CPRs may identify CPRs that are potentially useful in the particular clinical environment in which they are to be employed. This may involve developing an electronic query-based search strategy to identify candidate CPRs for further consideration. The identified literature-based CPRs provide the starting point for developing subsequent electronic CPRs deployed as decision support tools. One such systematic review of published literature identified CPRs specifically relevant to the family practice setting. Almost 800 published papers were identified, indicating the increasing level of research and interest in this area.32

---

**Figure 1. A Multistep Maturity Model for eCPRs**

![Diagram of the Multistep Maturity Model for eCPRs showing six levels: 1. Literature-Based CPRs, 2. Electronic Document-Based CPRs, 3. Electronic Computable Individual CPR Tools, 4. Service-Oriented Generalized CPR, 5. CPR with Terminology Services Integration, 6. Learning, Versionable CPR. Each level is connected by arrows indicating progression up the pyramid.](http://repository.academyhealth.org/egems/vol3/iss2/8)
Level 2: Electronic Document-Based CPRs


An improvement on the traditional literature-based CPR is to provide electronic document-based equivalents. These CPRs are not interoperable with other clinical systems in themselves but are documented as part of a collected, searchable register of rules in an electronic format with appropriate Web links to the original document-based sources. This is with a view to overcoming one of the initial difficulties in using CPRs by allowing more user-friendly searching and identification of appropriate CPRs for any presenting patient complaint, as shown in Figure 2. These search capabilities include searching based on the life cycle stage of the CPR, the clinical domain or condition targeted by the rule, and the clinical settings in which it is suitable for deployment.33

Figure 2. A Search Result from a Web-Based Register of CPRs (Example of Level 2 Electronic Document-Based CPR)
Level 3: Electronic Computable Individual CPR Tool

Interoperability Layers: *syntax – Stand-alone tool, potential to integrate within a single organization EHR.*

The majority of decision support tools implementing CPRs are at Level 3 of the model. Level 3 implements specific literature-based CPRs in decision support tools used at point of care in clinical practice. The representation of the rule is specific to the CPRs used and limited to use within the information systems in which they are deployed and tested. There may be some integration of the tools with another single organizational EHR and associated patient populations (a narrow CPR validation in practice). An important improvement is the wider dissemination of CPRs into clinical practice. The rule can be deployed electronically in a controlled clinical environment and made available to support subsequent validation and impact analysis efforts. This may take the form of randomized control trials testing the effectiveness of the electronic tool versus the performance of a control group without access to the tool.

Level 4: Service-Oriented Generalized CPR

Interoperability Layers: *syntax, transport – Interoperable and reusable with many different clinical applications within a single organization using open interoperability standards; lack terminology integration allowing access from other external systems that use different clinical coding schemes.*

The wider scale reuse of computable CPRs beyond their initial development environments may be achieved through a service-oriented architecture of CPR resources. This service-based approach has been increasingly deployed in DSSs as a means of promoting reuse of evidence and reducing development effort.

Broad validation of CPRs becomes possible when CPRs that were originally developed for use by an individual hospital department or family practice are reused as evidence to support wider dissemination, validation, and impact analysis in tools developed for other patient populations. This decouples the provision and querying of CPRs from the original deployment applications that use them. Evidence is accessed through widely used open standards from any development environment, thus supporting easier workflow integration.

The flexibility of such a service also depends on implementing a CPR model that captures computable structures common to all CPRs. This provides a model that can be used to deploy CPRs in a computable format that is accessible using open standards by third party tools via a Web service.

An example of a general model of CPRs that has been implemented is shown in Figure 3. The general model components captured in the model include the following:

- A presenting patient problem or reason for encounter (RFE);
- One or several differential diagnoses to consider associated with the RFE;
- A clinical prediction rule associated with a particular diagnosis;
- The CPR rule elements comprising of diagnostic cues and associated criteria to be checked, with a score that quantifies the significance of the rule element to the clinical outcome;
- A threshold-based score scheme that interprets the CPR possible score as risk bands with optional clinical action or recommendation to be carried out in response to the interpreted risk-based score bands;
- The demographic context of the derivation population from which the cue scores were derived for use by a CPR; and
- A standard code binding or clinical vocabulary term associated with each RFE, diagnosis, or diagnostic cue.
The general model of CPRs is then made available as a REST-based Web service that can be queried from any third-party development environment using open standards including XML, REST, and JSON.\textsuperscript{37-39} An example of a REST-based query from a decision support tool for a computable representation of the Alvarado Score is accessible using:

\texttt{http://localhost:8080/ClinicalEvidenceRESTService/interfaces/query/cprs/AlvaradoScore1.0}

The XML output using the model is generated and returned to the call decision support tool application as shown in Figure 4.
Figure 4. A Web Service-Based Call for Details of the Alvarado Score (Example of Level 4 Service-Oriented Generalized CPR)

```xml
</cpprelement>
+ <cpprelement>
  - <cpprelement id="query">
    - <cppElementCriteria>
      - <cppCriteriaScore>1</cppCriteriaScore>
      - <cppCriteriaPresent>true</cppCriteriaPresent>
    </cppElementCriteria>
    - <cppElementCue>
      + <codeBindings>
        <cueId>ReboundTenderness</cueId>
      </codeBindings>
    </cppElementCue>
    <cppElementNumber>5</cppElementNumber>
  </cpprelement>
  - <cpprelement id="query">
    - <cppElementCriteria>
      - <cppCriteriaScore>1</cppCriteriaScore>
      - <cppCriteriaPresent>true</cppCriteriaPresent>
    </cppElementCriteria>
    - <cppElementCue>
      + <codeBindings>
        <cueId>ElevatedTemperature</cueId>
      </codeBindings>
    </cppElementCue>
    <cppElementNumber>6</cppElementNumber>
  </cpprelement>
  - <cpprelement id="query">
    - <cppElementCriteria>
      - <cppCriteriaScore>2</cppCriteriaScore>
      - <cppCriteriaPresent>true</cppCriteriaPresent>
    </cppElementCriteria>
    - <cppElementCue>
      + <codeBindings>
        <codeId>WhiteBloodCellCount</codeId>
      </codeBindings>
    </cppElementCue>
    <cppElementNumber>7</cppElementNumber>
  </cpprelement>
+ <cpprelement>
  - <cppScoreScheme>
    <CPRScoreDecision>Discharge from hospital</CPRScoreDecision>
    <CPRScoreLevelEnd>4</CPRScoreLevelEnd>
    <CPRScoreLevelStart>1</CPRScoreLevelStart>
    <CPRScoreRisk>Low</CPRScoreRisk>
  </cppScoreScheme>
+ <cppScoreScheme>
+ <cppScoreScheme>
  <cppName>AlvaradoScore1_0</cppName>
</cppName>
</cppName>
</clinicalPredictionRule>
```
Level 5: CPRs with Terminology Services Integration

Interoperability Layers: semantics - Semantically interoperable with many different ICT applications across multiple organizations through addition of standard clinical code bindings.

The importance of integrating DSS tools into the wider clinical workflow has been highlighted as a key factor for their broader acceptance and implementation success. The capability for binding individual CPR-model terms with several clinical terminologies and vocabularies to support wider semantic interoperability and broader uptake of CPRs is crucial. This may be supported through providing the service-based CPR models in conjunction with clinical terminology or vocabulary services that enable terminology lookup, binding, and mapping of models to different vocabularies.

Integrating DSS tools with EHR systems based on coded patient data helps in identifying workflow related patient events that can be used as a contextual trigger for initiating diagnostic CPRs as a form of decision support. In addition the patient record data itself can then be utilized to provide patient demographics or patient historical data that may be used to contextualize CPR execution and selection of suitable scoring schemes based on the context of the particular patient.

As an example we have added code bindings supporting National Health Service (NHS) read codes widely used by EHR systems in the United Kingdom. Multiple code binding types can be added for each CPR cue to support other coding schemes used in other countries. An example of the output of one CPR cue element with code bindings for the “nausea” element of the Alvarado Score is shown in Figure 5. Where multiple patient codes may be suitable for triggering a cue, the “isPrimary” tag denotes the primary code and text to use for display in applications.

Level 6: Learning, Versionable CPR

Interoperability Layers: services - Interoperable with many different ICT applications across multiple organizations; capable of deriving CPRs electronically.

The development and continuous analysis of aggregated sources of electronic patient data to facilitate evidence generation and learning is a crucial part of the broader LHS vision. A number of existing aggregated sources of patient data are found at local and national levels. These contain large amounts of longitudinal population-health data suitable for data mining or statistical analysis with a view to deriving actionable knowledge.

The potential for using such data sources for data mining has been demonstrated in the TRANSFoRm project that utilized aggregated sources of European primary care data provided by the Transition project. An open source data-mining tool called KNIME was used to produce quantified association rule combinations describing the relationships identified between ICPC2 coded diagnostic cues, demographic variables, and diagnostic outcomes from the aggregated data sources. This process provided empirically quantified diagnostic associations using calculated likelihood ratios.

An example of a CPR construction tool creating a data-mined CPR for diagnosis of Urinary Tract Infection is shown in Figure 6. The tool presents data-mined evidence (left-hand side) through the Web service to construct versioned CPRs using the recognized formal CPR structure described in Level 3 (right-hand side). This allows for definition of normalized scoring schemes based on threshold approaches to decision-making. The score schemes risk levels and associated actions are defined through manual clinical review and interpretation of the general evidence and associated quality measures.
Figure 5. A Web Service-Based Call to Alvarado Score with Code Bindings (Example of Level 5 CPRs with Terminology Services Integration)
Saving the constructed CPR makes it available through the Web service, and it can be accessed using a standard Web-based call from other applications: [http://localhost:8080/ClinicalEvidenceRESTService/interfaces/query/cprs/DataMinedUTIRule](http://localhost:8080/ClinicalEvidenceRESTService/interfaces/query/cprs/DataMinedUTIRule).

The XML output of the call is shown in Figure 7.

Putting all the levels of the model together we can illustrate an architecture for electronic derivation of CPR evidence as shown in Figure 8.

Tracking of CPR versioning, change control, and usage will mean that in practice there should be at least two deployment environments such as a "development" and a "live" production CPR service. This can facilitate deployment of CPRs through the service for restricted narrow validation, and then promotion to wider usage for wider scale broad validation and impact analysis.
Figure 8. Summary of Electronic Derivation and Deployment of CPRs (Example of Level 6 Learning, Versionable CPR)
Clinical Workflow and User Interface Integration Considerations

The integration of eCPRs into existing electronic clinical care systems is crucial to their wider usage, but there are more factors to consider than simply the technical ones. User interface design considerations are also important to promoting the development and use of eCPRs more broadly in care settings. Studies have demonstrated the importance of consulting end-users regarding integration of eCPRs with existing systems. Another study that examined the deployment of EHR systems across the United Kingdom stressed the importance of “soft skills” such as training and multidisciplinary teams as being key to the uptake and usage of electronic clinical systems.

The type of CPR being developed should also be considered. CPRs may be related to diagnostic or prognostic outcomes. A diagnostic CPR estimates the probability or risk of the presence or absence of a disease at a fixed point in time, for a specific individual. A prognostic CPR is more complex, having an additional temporal aspect after the prognostic prediction has been made. It requires follow-up to see if a particular clinical event relating to the prognosis transpires at some defined subsequent point in time. The workflow implications are less complex for embedding diagnostic CPRs within EHRs or decision support applications, since the diagnostic CPR can be event driven and can trigger a recommendation made and recorded at a fixed point in time without the need for future follow-up.

The models described here focus on implementing diagnostic CPRs, and they are our primary examples. On that basis, interoperability considerations have been a core focus for development of this model before considering more complex time-dependent workflow integration.

Conclusion

The traditional focus on derivation and narrow validation of CPRs has severely limited their wider acceptance. The evolution and maturity model described here outlines a progression toward eCPRs achieving the vision of an LHS. The model provides an incremental framework consistent with the wider goals of the LHS and which demonstrates how we can consolidate work done by others in order to achieve this – using central repositories of CPR knowledge, accessible open standards, and generalizable models to avoid repetition of work. This is useful for developing more ambitious strategies to address limitations of the traditional CPR development life cycle, with a view to enabling wider implementation and acceptance of the benefits that intelligent use of CPRs can provide.

The model as presented here has limitations and can address only some of the issues with CPR development. For example, it does not address the design and type of clinical interventions to which any particular diagnostic CPR is best applied. It does not address what data elements may actually be available in a target EHR system to trigger event-driven use of our models. The wider availability of large volumes of aggregated data sources to support data mining approaches may be currently feasible only in limited cases.

It is time to look again at how we develop, disseminate, and test CPRs in clinical practice. The model is a starting point in promoting discussion about what a more dynamic CPR development process should look like.

Acknowledgements

This work was partially funded by the Health Research Board (HRB) of Ireland through the HRB Centre for Primary Care Research under Grant HRC/2007/1. TRANSFoRm is partially funded by the European Commission – DG INFSO (FP7 247787).
References

Findings from a Multi-Disciplinary Workshop Supported by the
Cornet R, de Keizer NF. The role of standardized data and
Ahmadian L, van Engen-Verheul M, Bakhshi-Raiez F, Peek N,
Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving
and Optimized retrieval of primary care clinical prediction
rules from MEDLINE to establish a Web-based register. J Clin
and developing an International Register of Clinical Prediction
rules from family practice to inform family practice; the learning
healthcare system case study on urinary tract infections. BMC
Family Practice. 2015;16(1):63.
A methodology for mining clinical data: experiences from TRANSFoRm project.
TRANSFoRm Project Homepage. Available from: http://www.
transformproject.eu.
Corrigan D, Hederman L, Khan H, Taweel A, Kostopoulou
O, Delaney B. An Ontology-Driven Approach to Clinical
Evidence Modelling Implementing Clinical Prediction Rules. In:
Moumtzoglou A, Kastania A, editors. E-Health Technologies
and Improving Patient Safety: Exploring Organizational
Corrigan D, Taweel A, Fahey T, Arvanitis T, Delaney B. Design
and Implementation of an Ontology for the Computable
Representation of Clinical Prediction Rules. EJBI. 2014;10(1):15-
23.
org/.
XML Technology Standard. Available from: http://www.w3.org/
standards/xml/.
Kawamoto K, Houllihan CA, Balas EA, Lobach DF. Improving
clinical practice using clinical decision support systems: a
systematic review of trials to identify features critical to
success2005.
Ahmadian L, van Engen-Verheul M, Bakhshi-Raiez F, Peek N,
Cornet R, de Keizer NF. The role of standardized data and
terminological systems in computerized clinical decision
support systems: literature review and survey. International
Design and Implementation of an Ontology for the Computable Representation of Clinical Prediction Rules

Derek Corrigan1, Adel Taweel2, Tom Fahey1, Theodoros Arvanitis3, Brendan Delaney2
1 HRB Centre for Primary Care Research, Royal College of Surgeons in Ireland, Ireland
2 Kings College London, United Kingdom
3 University of Warwick, United Kingdom

Abstract

Objective: A lack of acceptance has hindered the widespread adoption and implementation of clinical prediction rules (CPRs). The use of clinical decision support systems (CDSSs) has been advocated as one way of facilitating a broader dissemination and validation of CPRs. This requires computable models of clinical evidence based on open standards rather than closed proprietary content.

Methods: The on-going TRANSFoRm project has developed ontological models of CPRs suitable for providing CPR based decision support.

Results: This paper describes the design and implementation of a generic ontology model for the representation of computable CPRs. The conceptual validity and implementation of the ontology is discussed using an illustrative example of a CPR in the form of the Alvarado Score for acute appendicitis.

Conclusions: We demonstrate how the model is used to query the structure of this particular rule, providing a generic computable representation suitable for the representation of CPRs in general.

Keywords
Clinical prediction rules, ontology, clinical decision support

1 Introduction

Although many diverse examples of clinical prediction rules (CPRs) in primary care can be identified in research literature, their use has yet to gain widespread acceptance among clinicians [1, 2]. There are a number of valid concerns that influence why clinicians are reluctant to use them as part of their day-to-day clinical practice.

Despite the existence of an accepted development lifecycle for producing CPRs, many of them have traditionally focussed solely on the derivation phase of the CPR lifecycle [3]. Many derived CPRs are subject to poor or non-existent CPR validation and impact analysis. This lack of validation severely limits their perceived applicability to the same restricted patient populations defined in the original derivation research populations. Complications may arise when there are multiple rules derived by different researchers for any chosen clinical condition. As an example, a clinical condition such as Pulmonary Embolism has numerous variations of CPRs that may pertain to it [1]. This can lead to confusion and a lack of clarity about which CPR variations are the “correct” or “best” ones to use.

With some exceptions the format for dissemination of CPRs is largely literature based, putting an onus on clinicians to search literature for suitable CPRs [5]. This is compounded by the fact that literature based rules are by their nature static in content and do not provide for recording of versioned rule changes. This may have implications for the applicability of any particular CPR as changes take place over time in the demographics of the original rule derivation study population.

One suggested way of addressing these limitations is through development of clinical decision support systems (CDSSs) based on computable models of clinical evidence [6, 7, 8]. The ultimate vision is to provide for computable representations of CPRs that allow derivation, validation, dissemination, versioning and on-going revision from empirical sources of electronic primary care patient data. This can be complemented using extraction of patient cues...
and demographics from electronic health records (EHRs) as a trigger for initiating appropriate rule execution.

The TRANSFoRm project has developed computable ontological models of CPRs to support their electronic derivation, implementation and validation [9]. We describe the models and conceptual validity through implementation of a well studied CPR, the Alvarado score [10] [11]. We demonstrate how clinical questions are expressed as ontological queries for use by a CPR based CDSS currently being developed by the TRANSFoRm project.

2 Definition and application of Clinical Prediction Rules

2.1 CPR Definition

It is necessary to clearly define at the outset what we mean when we talk about using a clinical prediction rule. A CPR “is a clinical tool that quantifies the individual contributions that various components of the history, physical examination, and basic laboratory results make toward the diagnosis, prognosis, or likely response to treatment in a patient” [12] [13]. The formal characteristics of a CPR can be clearly identified based on this definition. Typically a CPR is derived from a statistical model and will be constructed and structured based on the following distinct parts:

- A clinical outcome that relates to a defined diagnostic, prognostic or treatment outcome associated with a selected clinical condition.
- A set of diagnostic cues and associated criteria that is indicative of the clinical outcome being assessed by the rule.
- A statistically derived scoring scheme that quantifies the relative contribution of each cue where present to the clinical outcome.
- A threshold based scoring scheme that defines relative clinical interpretations of risk categories for all possible scores for the rule.
- An optional decision indicating a clinical action in response to each risk category to be recommended based on each of the defined threshold scores.

2.2 Application of CPRs as part of a defined Diagnostic Strategy

In order to understand how CPRs may be potentially applied as a diagnostic tool in clinical practice it is useful to place their use in a broader diagnostic context. A clinician needs to formulate and consider the evidence for all possible differential diagnoses when a patient first presents with a particular clinical complaint. This is done by considering each differential diagnosis and can involve “ruling out” differentials based on the underlying diagnostic cues as presented by the patient. CPRs can provide a useful tool to assist with these potential “rule outs” using the results of applying suitable CPRs obtained to any particular patient case [13]. Their appropriate use can be applied as a tool to reduce the possibility of diagnostic error at the outset through consideration of possible differentials [12] [13]. As an example a patient presenting with abdominal pain who scores less than 4 on the Alvarado score, could indicate a potential “rule out” for acute appendicitis for that patient.

3 Model Development Methodology

The development of the formal models of clinical prediction rules described here followed a number of distinct steps subsequently described in detail:

- Clinical use case development.
- Functional requirements definition of the CPR model.
- Model design based on functional requirements.
- Model construction and clinical evidence population.
- Clinical use case implementation and validation.

3.1 Clinical Use Case Development

The models presented here provide the backend knowledgebase to be used as part of a broader piece of work currently in progress to develop a functional diagnostic decision support system as part of the TRANSFoRm project. The CDSS will consume and ask clinical questions of the models described here that provide the underlying knowledgebase. This CDSS tool will be deployed and used by primary care practitioners to assist them in formulating and quantifying differential diagnoses to consider for patients presenting with three defined diagnostic conditions. The use of electronic CPRs will be deployed as part of the diagnostic strategy for ruling out of potential differential diagnoses. The three primary care patient safety use cases will be used to test and validate the fully functional CDSS being developed by TRANSFoRm.

The selected patient safety use cases focus on potential diagnoses relating to patients presenting with the general complaints of chest pain, abdominal pain or dyspnoea. These were chosen for the cognitive challenge they present in primary care with potential for diagnostic error [14] [15]. Reviews of evidence based sources identified CPRs supporting selected diagnoses for these patient safety use cases [10] [16]. In total 41 clinical prediction
rules were identified relating to 20 diagnostic conditions relating to the three patient safety use cases. In this paper we describe the model representation of a single CPR called the Alvarado Score relating to a diagnosis of appendicitis for a patient presenting with abdominal pain.

3.2 Functional Requirements Definition of the CPR Model

In considering the model design requirements it is useful to first consider the functional requirements of any application that will use those developed models. The models described here will be ultimately queried by the TRANSFoRm CDSS. The CDSS will want to query particular diagnostic conditions, retrieve associated CPRs for any condition and query all of the constituent rule structures for any selected CPR. We have therefore defined our model requirements based on the different CPR related questions it needs to be able to answer. The functional requirements can be stated as clinical questions we wish to be able to ask of our CPR model. We identified the following questions as general functional requirements that we want to be able to answer using the finished CDSS tool:

- What are the differential diagnoses to consider for a selected patient reason for encounter (RFE)?
- What are the related CPRs associated with a selected diagnosis?
- What are the cues, criteria and associated scores of a selected CPR?
- What are the scoring interpretation schemes of a selected CPR?

Figure 1: Relationship of CPR ontology concepts.
• What are the population characteristics associated for application of a selected CPR?
• What is the clinical setting associated for application of a selected CPR?
• What are the supporting literature sources for a selected CPR?
• What is the current version number of a selected CPR?

3.3 Model Design based on Functional Requirements Definition

An ontology representation was chosen as the basis for the CPR model to support dissemination of CPRs using open standards that support a simple underlying data structure. Many methodologies have been proposed for design and development of ontologies [17]. The approach we have selected uses an application focussed approach where ontology requirements are expressed as “competency questions” that can then be used as a set of functional requirements to validate ontology completeness [18]. In our example, the functional requirements we have already defined can be considered to also define suitable ontology competency questions. If our ontology is designed correctly we should be able to express all our competency questions as formal ontology queries that generate correct clinical results with respect to our selected clinical use cases when executed (in this case representation of the Alvarado Score for appendicitis). Competency questions were deconstructed to identify the required formal ontology concepts and defined relationships that exist between them. The ontology concepts and relationships identified are shown in Figure 1. Although named inverse relationships exist for all relationships within the constructed ontology, we have only shown relationships in one direction in the diagram for clarity. For example the relationship ‘hasDifferentialDiagnosis’ has a corresponding inverse relationship called ‘isDifferentialDiagnosisOf’ that is not explicitly shown.

These core CPR ontology concepts are described in Table 1 along with examples of clinical instances and associated attributes of those concepts.

A fundamental requirement of the TRANSFoRm project is the appropriate use of standard clinical vocabularies, terminologies and classifications to add semantic meaning to any ontology terms being used through binding of vocabulary terms.

<table>
<thead>
<tr>
<th>Class Name and Description</th>
<th>Class Instance</th>
<th>Attribute Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>EvidenceRFE: AbdominalPainRFE</td>
<td>hasUMLSCode C0000737 hasICD10Code R10.0 hasReadCode XaA06 YaYkf</td>
<td></td>
</tr>
<tr>
<td>EvidenceDiagnosis: Appendicitis</td>
<td>hasUMLSCode C0003615 hasICD10Code K35 hasReadCode J20..Y30Di</td>
<td></td>
</tr>
<tr>
<td>ClinicalPredictionRule: AlvaradoScore1_0</td>
<td>hasRuleVersion 1_0 hasSupportingLiteratureURL <a href="http://www.biomedcentral.com/content/pdf/1741-7015-9-139.pdf">http://www.biomedcentral.com/content/pdf/1741-7015-9-139.pdf</a></td>
<td></td>
</tr>
<tr>
<td>ClinicalPredictionRuleElement: AlvaradoScoreElement1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EvidenceCue: Nausea</td>
<td>hasUMLSCode C0375548 hasICD10Code R11.0 hasReadCode X75qw.Y7Cjf</td>
<td></td>
</tr>
<tr>
<td>EvidenceCriteria: AlvaradoElementCriteria1</td>
<td>isPresent = True hasScoreInterpretation 1</td>
<td></td>
</tr>
<tr>
<td>ClinicalPredictionRuleScore: AlvaradoScoreLevel3</td>
<td>hasStartScore 7 hasEndScore 10 hasScoreInterpretation “Surgery”</td>
<td></td>
</tr>
<tr>
<td>EvidenceContext: Adult, Male, Ireland</td>
<td>hasAgeGreaterThan 17 hasISOCode 1 hasISOCode “IE”</td>
<td></td>
</tr>
<tr>
<td>EvidenceClinicalEnvironment: PrimaryCare</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A TRANSFoRm vocabulary service has been developed to allow runtime access to a number of vocabularies through defined web service interface methods [19]. The Unified Medical Language System [UMLS] has been used as a pivot terminology from which mappings have been provided to others including the International Classification of Primary Care Version 2 (ICPC2), SNOMED Clinical Terms, the International Classification of Diseases Version 10 (ICD10) and Read Codes [20, 21, 22, 23].

The CPR ontology model provides attributes (as shown in Table 1) to allow association of selected terminological codes to instances of the EvidenceRFE, EvidenceDiagnosis and EvidenceCue concepts. Multiple code system terms can be associated to any instance. At present these terms are manually entered into the ontology.

In order to facilitate CPR execution based on coded RFEs or diagnostic cues extracted from individual patient EHRs, future development will focus on integrating the ontology models with the TRANSFoRm vocabulary service. This will allow querying at runtime using only UMLS associations to pivot to the appropriate terminology implemented by the EHR data. This can also provide for coded ontology content to be represented and populated dynamically into the ontology through application of data mining techniques to electronic sources of coded primary care data.

### 3.4 Model Construction and Clinical Evidence Population

This constructed ontology design has been expressed using the ontology language/resource description framework (OWL/RDF) representation and implemented using the Protégé 4.1 ontology designer [24, 25, 26]. It is hosted using a Sesame triple store for query formulation, testing and future dynamic programmatic update of ontology content [27, 28]. The clinical content for the ontology was manually populated as instances of the ontology concepts to reflect the structure of the Alvarado score as described in literature [10, 11].

### 3.5 Ontology Metrics

The CPR ontology model is part of a larger clinical evidence ontology model that also supports the general representation of diagnostic knowledge. The knowledge-base metrics for the full ontology are:

- Number of ontology classes = 43
- Number ontology relationships = 101
- Data of ontology attributes = 48
- Number of ontology class instances = 505

### Table 2: Competency Questions 1-4 (from Table [1]) Expressed as SPARQL Queries with Associated Results.

<table>
<thead>
<tr>
<th>SPARQL (Protocol and RDF Query Language)</th>
<th>Query Result (Instance Relation Value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELECT ?DifferentialDiagnosis WHERE {?DifferentialDiagnosis isDifferentialDiagnosisOf AbdominalPainRFE .}</td>
<td>Appendicitis, BacterialEnteritis</td>
</tr>
<tr>
<td></td>
<td>ChronsDisease, CorPulmonale</td>
</tr>
<tr>
<td></td>
<td>EctopicPregnancy, Pyelonephritis</td>
</tr>
<tr>
<td></td>
<td>UrinaryTractInfection</td>
</tr>
<tr>
<td>SELECT ?CPR WHERE {?CPR isCprOf Appendicitis.}</td>
<td>AlvaradoScore1_0</td>
</tr>
<tr>
<td></td>
<td>MigrationOfPain hasScoreInterpretation 1</td>
</tr>
<tr>
<td></td>
<td>Anorexia isPresent true</td>
</tr>
<tr>
<td></td>
<td>Anorexia hasScoreInterpretation 1</td>
</tr>
<tr>
<td></td>
<td>Nausea isPresent true</td>
</tr>
<tr>
<td></td>
<td>Nausea hasScoreInterpretation 1</td>
</tr>
<tr>
<td></td>
<td>RightLowerQuadrantTenderness isPresent true</td>
</tr>
<tr>
<td></td>
<td>RightLowerQuadrantTenderness hasScoreInterpretation 2</td>
</tr>
<tr>
<td></td>
<td>ReboundPain isPresent true</td>
</tr>
<tr>
<td></td>
<td>ReboundPain hasScoreInterpretation 1</td>
</tr>
<tr>
<td></td>
<td>ElevatedTemperature isPresent true</td>
</tr>
<tr>
<td></td>
<td>ElevatedTemperature hasScoreInterpretation 1</td>
</tr>
<tr>
<td></td>
<td>Leucocytosis isPresent true</td>
</tr>
<tr>
<td></td>
<td>Leucocytosis hasScoreInterpretation 2</td>
</tr>
<tr>
<td></td>
<td>WhiteBloodCellShiftLeft isPresent true</td>
</tr>
<tr>
<td></td>
<td>WhiteBloodCellShiftLeft hasScoreInterpretation 1</td>
</tr>
<tr>
<td>SELECT ?ScoreElement ?Property ?Value WHERE {?ScoreElement isScoreSchemeOf AlvaradoScore1_0. ?ScoreElement ?Property ?Value.}</td>
<td>AlvaradoLevel1 hasStartScore 1</td>
</tr>
<tr>
<td></td>
<td>AlvaradoLevel4 hasScoreInterpretation &quot;Discharge&quot;</td>
</tr>
<tr>
<td></td>
<td>AlvaradoLevel1 hasScoreInterpretation &quot;Observation/Admission&quot;</td>
</tr>
<tr>
<td></td>
<td>AlvaradoLevel2 hasScoreInterpretation &quot;Observation/Admission&quot;</td>
</tr>
<tr>
<td></td>
<td>AlvaradoLevel2 hasStartScore 5</td>
</tr>
<tr>
<td></td>
<td>AlvaradoLevel6 hasScoreInterpretation &quot;Observation/Admission&quot;</td>
</tr>
<tr>
<td></td>
<td>AlvaradoLevel6 hasEndScore 10</td>
</tr>
<tr>
<td></td>
<td>AlvaradoLevel8 hasScoreInterpretation &quot;Surgery&quot;</td>
</tr>
<tr>
<td></td>
<td>AlvaradoLevel8 hasStartScore 7</td>
</tr>
<tr>
<td></td>
<td>AlvaradoLevel10 hasEndScore 10</td>
</tr>
</tbody>
</table>
4 Clinical Use Case Implementation and Validation

4.1 The Alvarado Score as a CPR example

A particular clinical example of a well studied CPR is the Alvarado Score which we will use as a clinical example to illustrate use of our models. This rule categorises the risk of patients having potential acute appendicitis based on the presence or absence of 8 diagnostic indicators. The risk of appendicitis is expressed as three score-based risk categories with associated recommended treatment options. This rule has been designed to be suitable for primary care and is based on the presence of diagnostic cues without the need for imaging [10]. Reviews have highlighted the importance of capturing the demographic context of the derivation study population. Clinical performance of the Alvarado score has been shown to vary in different populations depending on gender and age, performing best for adult males [11]. This demographic variability should be reflected in any model design.

Using the example of appendicitis and the Alvarado Score we identified the following questions as functional requirements that we want to able to answer using the finished CDSS tool:

- What are the differential diagnoses to consider for a reason for encounter (RFE) of abdominal pain?
- What are the CPRs associated with the differential diagnosis of appendicitis?
- What are the cues, criteria and associated scores of the Alvarado score?
- What are the scoring interpretation schemes of the Alvarado score?
- What are the population characteristics associated for application of the Alvarado score?
- What is the clinical setting associated for application of the Alvarado score?
- What are the supporting literature sources for the Alvarado score?
- What is the current version number of the Alvarado score?

4.2 Expression of CPR Model Queries

The competency questions previously defined as functional requirements were expressed as Protocol and RDF Query Language (SPARQL) ontology queries using the ontology concepts and relationships previously identified [28]. These queries were executed and results checked for consistency with respect to the clinical evidence sources used to populate the ontology. Queries and results are shown in Table 2 for four competency questions.

4.3 Development of Clinical Evidence Service

The evidence defined in the ontology has been made available to the TRANSFoRm CDSS through a REST based web service [29]. This allows the CDSS to access ontology resources through defined URL constructs that are linked to programmatically implemented SPARQL queries. The Sesame infrastructure provides a programmable API that can be used to programmatically connect to and query the ontology using SPARQL queries. The rest interface was developed using Java implementing the Jersey REST implementation [30].

System interoperability is supported by allowing query results to be returned to any third party consumer tool in a number of supported data formats including XML, JSON and plain text responses. In addition, the Sesame infrastructure also provides its own REST based interface that can be used directly to execute SPARQL queries to return responses in native RDF data formats. The components of the evidence service are shown in Figure 2.

![Figure 2: Evidence service components.](image-url)
4.4 Implementation of a Diagnostic Strategy using Evidence Service Calls

We previously referred to the role of CPRs as part of a broader diagnostic strategy to “rule out” a potential diagnosis. The steps to be implemented would require:

- Obtaining the list of supported patient RFEs
- Obtaining a list of differential diagnoses to consider based on a presenting patient RFE
- Obtaining a list of the CPRs available for any differential diagnosis associated with the RFE
- Obtaining the cues, criteria and scores for any chosen CPR to apply
- Obtaining the scoring scheme and decisions for any chosen CPR
- Execution of the CPR based on a comparison to the patient cues provided to determine if a “rule out” may be appropriate

Using the example of a patient presenting with abdominal pain and an investigation of possible appendicitis, the clinical evidence service can be used to implement these steps using the following series of REST based calls to present results as XML, JSON, plain text or RDF data formats:

```plaintext
../ClinicalEvidenceRESTService/interfaces/query/rfes
../ClinicalEvidenceRESTService/interfaces/query/differentials/AbdominalPainLocalisedOtherRFE
```

![XML for Alvarado Score](http://localhost:8080/ClinicalEvidenceRESTService/interfaces/queries/cprs/AlvaradoScore1_0)

Figure 3: XML generated from evidence service to provide criteria for the Alvarado Score rule.
Corrigan et al. – Design and Implementation of an Ontology for the Computable Representation CPRs

The workflow content of these clinical decisions could be owned right within the ontology e.g. CPRClinicalDecision. It could be possible though the ontology e.g. “Surgery”. It could be possible though the model to represent 41 clinical prediction rules relating to 20 diagnoses including the Finnish Diabetes Risk Score [32], the Edwards Score [33] (tuberculosis) and the Little Symptom rule [34] (urinary tract infection).

In considering how this model relates to other initiatives to represent electronic clinical guidelines it is important to consider the original definition of a CPR previously provided. Each CPR is defined to be a discrete independent clinical tool to be used in its own right with respect to a particular patient. They do not attempt to define a complex clinical workflow or series of clinical steps to be implemented. From this point of view they are potentially useful tools to support decision-making in primary care where time pressures apply to consultations with each patient. As such, they could be considered to be either stand alone tools or are analogous to decision points found in more complex electronic guidelines that do define computable workflows, such as Guideline Interchange Format (GLIF) or the Guideline Elements Model (GEM) [35, 36].

The previous definition of a CPR also allows for an optional clinical decision or action to be taken based on the score outcome of the rule (sometimes then referred to as a Clinical Decision Rule). It was considered to be out of the scope of this work to represent these clinical decisions as computable entities in their own right and they have been treated as informational textual descriptions in the ontology e.g. “Surgery”. It could be possible though to represent these decisions as separate concepts in their own right within the ontology e.g. CPRClinicalDecision. The workflow content of these clinical decisions could be modelled separately as GLIF or GEM based guidelines with an appropriate reference or link from our ontology concepts.

There are limitations to this work because the TRANSFoRm project as a whole is still a work in progress. The future development of the clinical decision support system that consumes our evidence service will be necessary to do a full clinical validation of the models that we propose here. What we have presented here is a conceptual validation of the ontology structure and the implementation in a way that supports system interoperability (through recognised data representation standards such as XML, JSON and RDF) along with semantic interoperability (through the use of the TRANSFoRm vocabulary service).

5 Discussion

A core requirement for the development of the CPR model was that it be a generalisable representation of the common structure of CPRs and not just suitable for the representation of specific examples of rules as found in literature. The efficacy of using CPRs as tools to be deployed in decision support systems has been shown to be effective but focussed on implementing specific instances of CPRs rather than supporting their more general usage through a service based knowledgebase [31]. We have used the model to represent 41 clinical prediction rules relating to 20 diagnoses including the Finnish Diabetes Risk Score [32], the Edwards Score [33] (tuberculosis) and the Little Symptom rule [34] (urinary tract infection).

In considering how this model relates to other initiatives to represent electronic clinical guidelines it is important to consider the original definition of a CPR previously provided. Each CPR is defined to be a discrete independent clinical tool to be used in its own right with respect to a particular patient. They do not attempt to define a complex clinical workflow or series of clinical steps to be implemented. From this point of view they are potentially useful tools to support decision-making in primary care where time pressures apply to consultations with each patient. As such, they could be considered to be either stand alone tools or are analogous to decision points found in more complex electronic guidelines that do define computable workflows, such as Guideline Interchange Format (GLIF) or the Guideline Elements Model (GEM) [35, 36].

The previous definition of a CPR also allows for an optional clinical decision or action to be taken based on the score outcome of the rule (sometimes then referred to as a Clinical Decision Rule). It was considered to be out of the scope of this work to represent these clinical decisions as computable entities in their own right and they have been treated as informational textual descriptions in the ontology e.g. “Surgery”. It could be possible though to represent these decisions as separate concepts in their own right within the ontology e.g. CPRClinicalDecision. The workflow content of these clinical decisions could be modelled separately as GLIF or GEM based guidelines with an appropriate reference or link from our ontology concepts.

There are limitations to this work because the TRANSFoRm project as a whole is still a work in progress. The future development of the clinical decision support system that consumes our evidence service will be necessary to do a full clinical validation of the models that we propose here. What we have presented here is a conceptual validation of the ontology structure and the implementation in a way that supports system interoperability (through recognised data representation standards such as XML, JSON and RDF) along with semantic interoperability (through the use of the TRANSFoRm vocabulary service).

6 Conclusion

The research described in this paper can encourage the wider use and acceptance of clinical prediction rules by clinicians in three ways; by making CPRs more accessible and searchable than literature equivalents; through provision of a computable representation that allows for development of versioned rules from data mined sources of aggregated primary care data that are more sensitive to clinicians own patient populations; through provision of a web service allowing the deployment of CPRs as part of third party decision support tools linked to EHRs to facilitate easier use and execution.

Acknowledgements

TRANSFoRm is partially funded by the European Commission – DG INFSO (FP7 247787). This research was partly supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy’s and St Thomas’ NHS Foundation Trust and King’s College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

This work was partially funded by the Health Research Board (HRB) of Ireland through the HRB Centre for Primary Care Research under Grant HRC/2007/1.

References


[9] TRANSFoRm Project Homepage [Internet]. TRANSFoRm project [cited 31 July 2013]. Available from http://www.transformproject.eu


An ontology driven clinical evidence service providing diagnostic decision support in family practice

Derek Corrigan, MSc
HRB Centre for Primary Care Research,
Royal College of Surgeons in Ireland,
Dublin,
Ireland

Abstract
Formulation of a working diagnostic hypothesis in family practice requires consideration of many differential diagnoses associated with any presenting patient complaint. There follows a process of refinement of the differentials to consider, through ruling in or out each candidate differential based on the confirmed presence or absence of diagnostic cues elicited during patient consultation. The patient safety implications of diagnostic error are potentially severe for patient and clinician. This paper describes a clinical evidence service supporting this diagnostic process. It allows decision support consumers to provide coded evidence-based recommendations to assist with diagnostic hypothesis formulation, integrated with an EHR in primary care. The solution implements ontology models of evidence accessible to consumers as a web service using open source components and standards. An implementation example is described that consumes the service to drive a diagnostic decision support tool developed for the TRANSFoRm project.

Introduction and Background
The nature of family practice requires familiarity with a wider range of clinical conditions than specialists working in secondary care. The diagnostic process in family practice requires formulation of a working diagnosis based on the primary presenting patient complaint or reason for encounter (RFE) (1). Consideration is given to each candidate differential diagnosis with a view to ruling it in or out based on the confirmed patient diagnostic cues identified through consultation (2). This process can fail and diagnostic error has been shown to be a major threat to patient safety in the family practice setting (3-4). The knowledge base available to any clinician may be limited by their own case experience or, in the case of rare or unfamiliar conditions, distorted through cognitive bias. This may distort the initial formulation of differentials to consider which has been shown to be a crucial first step in the diagnostic process (5). This paper describes the design, development and implementation of a clinical evidence service for the TRANSFoRm project (6). The service supports the diagnostic process in family practice by allowing querying and refinement of coded clinical evidence supporting candidate differentials to consider based on submission of the patient presenting RFE and consultation diagnostic cues.

Methods
Evidence Service Design
The clinical concepts to support a diagnostic process have been modelled as an ontology of clinical evidence (figure 1). This allows representation of the relationships between a presenting patient reason for encounter and the associated candidate differential diagnoses to consider. The evidence relating to any particular diagnosis is captured as associated diagnostic cues, of which there are cue sub-concepts to represent clinician observed signs, patient reported symptoms, risk factors and clinical tests. The ontology design methodology used is based on the design practices advocated by the work of Gruninger and Fox (7). Semantic interoperability support is provided for by the ‘code binding’ concept, separate and independent of the clinical concepts. This is used to associate potentially many different clinical terminology codes for any single ontology RFE, cue or diagnosis.
Specifically we have supported interoperability with a single EHR vendor in the UK using NHS read codes version 3 (8). Localisation support to allow easier searching by a third party consumer for ontology terms using locally defined synonyms is provided for by the ‘synonym’ concept. An example of concepts and associated instances (in red) of the diagnostic cue concept for a patient history of irritable bowel syndrome, with an associated NHS read code ‘14CF.00’ and local synonym ‘HO IBS’ is shown in figure 2. The ontology can support other coding schemes including ICPC2, ICD10, SNOMED and UMLS (9-11).

![Image of core evidence ontology concepts](image1.png)

**Figure 1:** Core evidence ontology concepts.

![Image of cue ontology concept instance for ‘history of irritable bowel syndrome’](image2.png)

**Figure 2:** Example of cue ontology concept instance for ‘history of irritable bowel syndrome’.

### Evidence Service Implementation

The clinical evidence service consists of three implementation layers. The ontology is implemented as an OWL ontology using Protégé version 4.3 and hosted on a Sesame triple store (12-14).

![Image of evidence service implementation technologies](image3.png)

**Figure 3:** Evidence service implementation technologies.

Using structured evidence service endpoints we can access any ontology content with results returned as XML (default), JSON or RDF formats (16-18). The REST query to access the differentials to consider for a patient presenting with abdominal pain for example is:
To access the cues supporting diagnosis of urinary tract infection (output shown in figure 4) the query is:

http://phaedrus.scss.tcd.ie/munnellg/ClinicalEvidenceRESTService/interfaces/query/differentials/cues/UrinaryTractInfection

Sesame also provides flexibility beyond the defined endpoints by providing functionality to process custom ad-hoc SPARQL queries executed directly against its own accessible web service interface.

The client layer provides a client side library used to handle exchange of patient data between the third party consumer with appropriate calls sent to the backend evidence service. The client accepts patient data in the form of a XML patient evidence set describing the patient RFE, demographics and the underlying cues confirmed through consultation with the patient (figure 5). The evidence service returns recommendations in the form of a dynamically updated ranked list of differentials to consider by keeping a cue count for each differential under consideration. This list is based on the presenting RFE and ordered in descending cue count based on the number of patient cues confirmed present for each differential along with the supporting underlying evidence cues for each diagnosis. An
interactive and iterative diagnostic conversation can take place between the third party consumer as presence or absence of patient cues are confirmed, appropriate patient contextualised REST queries are executed and the re-ranked diagnosis list is supplied to the consumer tool.

**Results**
The evidence service implements diagnostic content from systematic review of evidence based sources that identified appropriate diagnostic cues for seventy eight diagnostic conditions relating to three presenting patient complaints: abdominal pain, chest pain and dyspnoea. The TRANSFoRm project has used this to provide ontology driven prompting and recording of coded patient diagnostic cues from a separately developed diagnostic decision support tool embedded and interoperable with an EHR in family practice (the Vision 3 EHR). This allows bottom-up input of observed patient cues independent of associated diagnosis (left window) or top-down drilling into and selection of evidence cues supporting specific diagnoses (right window). A dynamically updated cue count is maintained for each differential diagnosis indicating the number of evidence cues that are confirmed as present based on the patient cues elicited. As each patient cue is selected the cue counts are recalculated. The differentials are re-ranked in descending order based on the cue count. In addition each diagnosis has a prevalence category assigned to it (common, uncommon and rare). Where differentials have the same number of cues present they are ordered by highest prevalence first. These are used to dynamically rank potential differential diagnoses to consider (most likely at top) based on the patient presenting RFE along with the evidence supporting each diagnosis under consideration (figure 6). Upon exiting the tool a working diagnosis can be confirmed and the coded evidence cues and current working diagnosis can be saved back and recorded for future reference in the patient EHR.

![Figure 6: The diagnostic decision support window accessible from the patient EHR record shown in the background](image)

The tool is separately undergoing evaluation of diagnostic accuracy and ease of use by researchers at Kings College London. A sample of 32 UK family practitioners has been trained to use the tool prior to use. Diagnostic performance is compared between family practitioners with and without access to the tool. Performance is tested against predefined diagnostic scenarios represented by actors in simulated family practice encounters. Ease of use is assessed using a tool usability questionnaire.
Discussion
The implementation of the clinical evidence service can be contrasted with other comparable approaches (19). The unique features of this work include the provision of an evidence base provided through an open service-oriented architecture combined with an ontology model that is independent of any specific diagnostic condition. The provision of an ontology model driven by the concept of the RFE allows for a combination of top-down or bottom-up diagnostic reasoning making it particularly suited to the requirements of the diagnostic process in family practice. The hosting of the ontology on a triple-store platform is flexible enough to provide for future population of content programmatically from dynamically generated clinical content using identification of diagnostic associations from aggregated sources of coded EHR data. The TRANSFoRm project has already developed a data mining tool to allow for generation of quantified empirical evidence based on association rules that can be imported into the evidence service. Future work will assess the feasibility of using data mined evidence to rank differentials under consideration using Bayesian methods.

Acknowledgements
The TRANSFoRm project is partially funded by the European Commission - DG INFSO (FP7 247787).

References
Evidence-based rules from family practice to inform family practice; the learning healthcare system case study on urinary tract infections

Jean K Soler¹, Derek Corrigan²*, Przemyslaw Kazienko³, Tomasz Kajdanowicz³, Roxana Danger⁴, Marcin Kulisiewicz³ and Brendan Delaney⁵

Abstract

Background: Analysis of encounter data relevant to the diagnostic process sourced from routine electronic medical record (EMR) databases represents a classic example of the concept of a learning healthcare system (LHS). By collecting International Classification of Primary Care (ICPC) coded EMR data as part of the Transition Project from Dutch and Maltese databases (using the EMR TransHIS), data mining algorithms can empirically quantify the relationships of all presenting reasons for encounter (RfEs) and recorded diagnostic outcomes. We have specifically looked at new episodes of care (EoC) for two urinary system infections: simple urinary tract infection (UTI, ICPC code: U71) and pyelonephritis (ICPC code: U70).

Methods: Participating family doctors (FDs) recorded details of all their patient contacts in an EoC structure using the ICPC, including RfEs presented by the patient, and the FDs’ diagnostic labels. The relationships between RfEs and episode titles were studied using probabilistic and data mining methods as part of the TRANSFoRm project.

Results: The Dutch data indicated that the presence of RfEs “Cystitis/Urinary Tract Infection”, “Dysuria”, “Fear of UTI”, “Urinary frequency/urgency”, “Haematuria”, “Urine symptom/complaint, other” are all strong, reliable, predictors for the diagnosis “Cystitis/Urinary Tract Infection”. The Maltese data indicated that the presence of RfEs “Dysuria”, “Urinary frequency/urgency”, “Haematuria” are all strong, reliable, predictors for the diagnosis “Cystitis/Urinary Tract Infection”. The Dutch data indicated that the presence of RfEs “Flank/axilla symptom/complaint”, “Dysuria”, “Fever”, “Cystitis/Urinary Tract Infection”, “Abdominal pain/cramps general” are all strong, reliable, predictors for the diagnosis “Pyelonephritis”. The Maltese data did not present any clinically and statistically significant predictors for pyelonephritis.

Conclusions: We describe clinically and statistically significant diagnostic associations observed between UTIs and pyelonephritis presenting as a new problem in family practice, and all associated RfEs, and demonstrate that the significant diagnostic cues obtained are consistent with the literature. We conclude that it is possible to generate clinically meaningful diagnostic evidence from electronic sources of patient data.

Keywords: Learning healthcare system, Data-mining, international classification of primary care, Diagnosis, Reason for encounter, Urinary tract infection, Pyelonephritis, Transform, Transition project, Electronic patient record

* Correspondence: derekcorrigan@rcsi.ie

© 2015 Soler et al; licensee BioMed Central. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Background

The process of diagnosis in family medicine (FM, synonymous with general practice) can be informed and enhanced by evidence emerging from data collected in routine clinical practice. The analysis of data on the elements of the encounter relevant to the diagnostic process sourced from routine electronic medical record (EMR) databases represents a classic example of the concept of a learning healthcare system (LHS).

The International Classification of Primary Care (ICPC) acts as an ordering principle for FM data, allowing for direct international comparisons, and has the appropriate granularity for primary care data aggregation and analysis [1-5]. In the Transition Project, such ICPC data have been collected with EMRs in the Netherlands, Japan, Poland, Malta, Serbia, and other countries from the daily practice of a cohort of family doctors (FDs) using a similar methodology over time (one to eleven years) [4,6]. Such data have been used in the TRANSfoRm project to develop a diagnostic decision support system (DDSS) for FM, and the data have been recently published online as a repository of diagnostic association rules which are free to use to support family doctors’ (FDs) diagnostic processing [7,8].

The use of ICPC to study the epidemiology of FM has the advantage of allowing precise capture of reason for encounter (RfE) data, often ignored in FM research, and this allows further important perspectives into the process of diagnosis in FM [3,5,9-12].

This paper aims to exemplify the use of FM data to support diagnostic decisions in routine practice by analysing all possible associations between all the presenting RfEs in the Dutch and Maltese Transition Project databases (using the EMR TransHIS) and new episodes of care (EoC) for urinary system infections: simple urinary tract infection (UTI, ICPC code: U71) and pyelonephritis (ICPC code: U70).

The research question for this study is: “What are the quantitative relationships between reasons for encounter and the diagnoses ‘UTI’ and ‘pyelonephritis’ (episode titles) within new episodes of care in routine family practice in practice populations from Malta and the Netherlands and how do these match the data already published in the literature?”.

Methods

The public-domain EMR TransHIS, designed for use with ICPC, was used to collect data from participating FDs who recorded details of all their patient contacts in an episode of care (EoC) structure using ICPC. The study did not involve the collection of new data. Ethical approval was applied for locally, when appropriate, for individual studies based on these data in the Netherlands, Serbia and Malta. Reasons for encounter (RfEs) presented by the patient, all FD interventions, and the diagnostic labels recorded for each encounter were classified as recommended with ICPC (ICPC-2-E in Malta and Serbia, ICPC-1 in the Netherlands). All encounter data (face to face encounters in the office and at home, telephone consultations, repeat prescriptions, etc.) were analysed in an EoC structure to obtain complete data on incidence and prevalence, including patients presenting for a repeat prescription only [4-6,9-12].

An EoC is defined as a health problem from its first presentation by the patient to the FD, until the completion of the last encounter for it. It encompasses all contact elements related to that health problem. Its name (i.e. the diagnostic label of the EoC) may be modified over time, and in this article we refer to it as the episode title [2]. The last diagnosis made during an EoC is the current episode title. In this study, we focus on only two episode titles: urinary tract infection and pyelonephritis.

The RfE(s) is defined as an agreed statement of the reason(s) why a person enters the health care system, representing the demand for care by that person. The RfE should be recognized by the patient as an acceptable description of the demand for care [2]. FDs recording data for the Transition Project were trained to record RfEs according to the definitions above and the recommendations in the ICPC book, reflecting the patient’s symptoms and requests as expressed. Symptoms elicited during history-taking (i.e. the history of the presenting complaint) were recorded in a separate cell in the EMR TransHIS, but were not included in the analyses in this study [5,11].

The two databases each encompass a defined time period: an average of 9,896 patients and 43,577 patient-years of observation over 5 years in Malta (2001–2005), and 15,318 patients and 158,370 patient-years over 11 years in the Netherlands (1995–2005). The practice populations in the Netherlands represent the registered patients, whilst the population in Malta represents the patients consulting over a five year period [5,9]. These databases are available in the public domain (www.transitieproject.nl; www.mipc.org.mt).

The relationships between RfEs and episode titles were studied using Bayesian probabilistic methods. According to Bayes’ Theorem, the post-test (posterior) odds of an event (i.e. a specific diagnosis being made) are equivalent to the pre-test odds multiplied by the likelihood ratio (LR) [5,10,11]. The LR values were calculated for a problem presenting for the first time at the beginning of a new EoC.

The LR is a mathematical, quantitative expression of the extent to which a symptom increases the probability of a given diagnosis. The positive LR (LR+) for the existence of the symptom is the odds that it will exist in a patient with the disease (relevant to diagnosis), in contrast to a patient without the disease. The negative LR (LR-) for absence of the symptom is the odds that the test will
be negative in a patient with the disease, contrasted with a patient without the disease. We aggregated or pooled likelihood ratios across practices, as we have done in our previous studies [11,12].

It is possible to analyse such relationships between all possible combinations of episode titles and RfEs, using the TransHIS databases. Such an analysis has been performed and is presented on-line [8]. The website allows browsing, filtering, sorting and commenting of the results (association rules) of a data mining analysis platform that has been implemented to generate actionable clinical knowledge from electronic sources of coded primary care data. The user can filter the rules according to RfE, diagnostic cue (Anamnesis), diagnoses, sex and age groups as well as various quantitative measures. An open source data analysis tool, the Konstanz Information Miner (KNIME) has been used to define workflows that pre-process the TransHIS record data and derive association rules based on ICPC2 codes [13]. These rules identify all possible combinations of RfE, diagnostic cues and demographic variables (antecedent variables) that are linked with a recorded diagnostic outcome (consequent variable). The patient records loaded into KNIME consisted of only the first patient encounter relating to each new EoC for any patient. After cleaning (first encounter only from new episodes) 393,169 patient encounters were loaded into KNIME: 55,821 for Malta and 337,348 for the Netherlands. In total, 542,739 association rules were extracted from the data: 61,563 for Malta, 191,883 for the Netherlands, and 289,293 for both populations combined.

The distinct steps implemented in the data mining process required:

- derivation of association rules linking RfE, diagnostic cues and demographics to the recorded diagnoses made during the first encounter of a new episode of care
- calculation of association rule quality measures to determine the relative strength of each rule association derived, among others LR+, LR-
- filtering of association rules to allow selection of ‘high-quality’ association rules
- clinical review of selected rules to assess clinical validity of rules with respect to wider clinical body of evidence.

The analysis presented here was limited to two selected diagnoses, urinary tract infection (UTI) and pyelonephritis (U70), for practical reasons. The minimum level of clinical significance for a LR was arbitrarily taken as representing a standardised difference of at least 0.10 (10%) [14]. Cut-off levels of 2 for the LR+ of a positive association, and <0.5 for the LR- of a negative association, were thus taken as minimum thresholds for clinical significance. LRs outside these limits were considered not clinically significant. On the other hand, LRs outside a second arbitrary threshold (LR+ >8, LR- <0.2) were considered to indicate a strong diagnostic association, and indicated as such in our conclusions [11,12]. Furthermore, as above, LRs which were not at least as large as their 95% confidence level (CI) were considered unreliable [11,12,14]. LRs based on cells with very small numbers were ignored. These criteria adjust for the increased chance of describing spurious associations due to the large number of repeated statistical tests in our analytic process, and also for errors in under-estimation of variance due to the effect of clustering [11,12,14].

Results

The raw data output from the analysis applied using KNIME is provided in Additional file 1 and summarised in the form of tables in Additional files 2 and 3. These tables show the diagnostic associations identified for “UTI” and “Pyelonephritis” respectively. The positive likelihood ratios for all associated RfEs and the episode title “UTI” or “Pyelonephritis” in the two populations are listed. LRs are highlighted according to its value (clinical significance) and reliability (95% CI). Strong predictors (LR+ >8 or LR- <0.2, CI width smaller than or equal to the LR itself) are shown in red highlight. Weak predictors (LR+ >2-8, LR- 0.2-0.4, small CI) are shown in green highlight. LRs with a wide CI (larger than the observation itself) or which are not clinically significant (LR+ < =2, LR- > =0.5) or have a CI which includes unity are not highlighted.

The Dutch data and the combined dataset for “UTI” (Additional file 2) indicated that the presence of RfEs “Cystitis/Urinary Tract Infection”, “Dysuria”, “Fear of UTI”, “Urinary frequency/urgency”, “Haematuria”, “Urinary symptom/complaint, other” are all strong and reliable predictors for the diagnosis “Cystitis/Urinary Tract Infection”. The RfEs “Incontinence urine,” “Urination problems, other”, “Abdominal pain localised, other”, “Flank/axilla symptom/complaint” are all reliable, but less strong predictors for the diagnosis “Cystitis/Urinary Tract Infection”. In the Dutch data the presence of RfEs “Vaginal symptom/complaint” or “Vaginal discharge” are strong but unreliable predictors to exclude a diagnosis of “Cystitis/Urinary Tract Infection”. The combined dataset indicated that “Vaginal symptom/complaint” was no longer a predictor for excluding a diagnosis of “Cystitis/Urinary Tract Infection”.

The Maltese data for “UTI” (Additional file 2) indicated that the presence of RfEs “Dysuria”, “Urinary frequency/urgency”, “Haematuria” are all strong, reliable, predictors for the diagnosis “Cystitis/Urinary Tract Infection”. The RfE “Abdominal pain localised, other” is a less strong but reliable predictor for the diagnosis “Cystitis/Urinary Tract Infection”.
In Additional file 3, the diagnostic associations for “Pyelonephritis” are analysed. The Dutch data indicated that the presence of RfEs “Flank/axilla symptom/complaint”, “Dysuria”, “Fever”, “Cystitis/Urinary Tract Infection”, “Abdominal pain/cramps general” are all strong, reliable, predictors for the diagnosis “Pyelonephritis”. The RfEs “Vomiting,” “Back symptom/complaint”, “Urinary frequency/urgency”, “Nausea”, “Abdominal pain localised, other”, “Low back symptom/complaint” are all less strong, but reliable predictors for the diagnosis “Pyelonephritis”. The combined dataset resulted in a number of weak predictors from the Dutch dataset becoming insignificant predictors. This loss of significance is due to the smaller number of cases of pyelonephritis combined from the Malta dataset.

The Maltese data set did not present any clinically and statistically significant predictors for pyelonephritis.

Discussion
Principal findings
This is a study of the clinical interpretation of two common symptom diagnoses, “Cystitis/Urinary Tract Infection” and “Pyelonephritis”, in routine family practice in two practice populations, Malta and the Netherlands as well as both combined. The data collected with ICPC were used to analyse the RfE associations between these two diagnoses made during the first encounter of an EoC starting with their presentation to the FD. A number of positive and negative diagnostic associations were found between these two RfEs and a number of episode titles. These associations were found to have different strengths of effect and differing precision of the effect estimate. However, a number of diagnostic associations were found to be similar across the two populations. A larger population would have given more precise LR estimates, and would likely have demonstrated even more congruence between these diagnostic associations.

Implications of the findings
This study presents diagnostic associations from the perspective of the RfE, making it particularly useful to clinicians dealing with diagnostic challenges in the form of a newly presenting symptom in their daily practice. There were more similarities than differences in the diagnostic associations between RfEs and episode titles across populations, especially evidenced by the more frequent observations with narrower CIs.

Comparisons with the literature
A key objective of this analysis was to compare for consistency of the clinical associations generated from our analysis with previous high quality studies of clinical evidence relating to the two diagnostic conditions. As such high quality evidence based reviews or guidelines of clinical evidence supporting Urinary conditions was chosen for comparison (Bent et al. 2002, SIGN UTI Guidelines 2012, European urology Guidelines 2013) [15-17]. A high level summary from the SIGN guidelines gives the following symptom based definitions of cystitis and pyelonephritis:

- UTI - “evidence of urinary tract infection with symptoms suggestive of cystitis (dysuria or frequency without fever, chills or back pain)”
- Pyelonephritis - “evidence of urinary tract infection with symptoms suggestive of pyelonephritis (loin pain, flank tenderness, fever, rigors or other manifestations of systemic inflammatory response)” [16].

The European Urology guidelines define cystitis symptoms as “Dysuria, frequency, urgency, pain or bladder tenderness”. These symptoms progress to pyelonephritis with additional symptoms of “Fever, Flank pain, Nausea, vomiting” [17].

The identified predictors from our analysis compare favourably with both the cystitis and pyelonephritis definitions. Our analysis indicated similar predictors in the form of urinary frequency, haematuria and dysuria from both population data sets for cystitis. Other weaker predictors are consistent including abdominal pain or flank pain. Predictors for pyelonephritis such as fever, flank/back pain, nausea and vomiting were also consistent with literature. In the Netherlands dataset, self-labelling by patients was also shown as a strong predictor for UTI. The presence of vaginal discharge was not quite strong enough to be considered a definitive excluding factor. Unlike the JAMA review, no association with fever was found for cystitis and this is consistent with later SIGN and European Urology guidelines which indicate this should be considered indicative of progression to pyelonephritis [15]. However we could not confirm any negative relationships between the presence of the symptoms “fever,” “chills” or “back pain” and a diagnosis of “Cystitis/Urinary Tract Infection.” Our analysis also highlighted cystitis itself as a significant predictor for pyelonephritis indicating the relationship and progression of these conditions into each other.

The JAMA review with quantified likelihood ratios for specific cues concluded that “specific combinations of symptoms (e.g. dysuria and frequency without vaginal discharge) raise the probability of UTI to more than 90%, effectively ruling in the diagnosis based on history alone” [15]. In our analysis dysuria with frequency was found to be the single biggest LR for a combination of cues and is consistent with the JAMA conclusions. Our calculated likelihood ratios were generally stronger than those from JAMA reviews which reflects firstly the larger volumes of data analysed in this study which did not
pre-select patients with the index conditions, and secondly the effect of the lower prior probability with the earlier presentation of illness in primary care, as against emergency and secondary care (and consequently a higher positive LR). The clinical data published in the literature rarely include LRs based on primary care data, and further comparisons were therefore not possible, although desirable.

Comparison between populations
The number of associations and their relative strengths were found to improve with analysis of larger volumes of data as shown by the relative comparison of generated associations from Netherlands and Malta. The smaller volume of Malta data tended to generated LRs that had wider confidence intervals. The prevalence of the condition has also shown to be important in requiring larger volumes of data as shown by the lack of predictors identified for the rarer pyelonephritis found in Malta data. The key cystitis indicators from Malta are consistent with the Netherlands data.

The relative lack of symptoms-oriented research into the diagnostic process in primary care makes finding comparable literature challenging. Most studies of diagnostic associations have been performed in datasets which are not exclusively or mainly from primary care. Additionally, most study a disease-label diagnosis and its associations with symptoms and test results as predictors, and not the other way around. In that sense, the diagnostic associations we have found may be more acceptable to and useful for clinicians. Furthermore, the congruency (and often statistical consistency) of diagnostic associations between these populations, and especially the fact that most of them are in the same direction from unity, sustain our confidence in their validity.

The results for the combined data (Malta and the Netherlands together) are heavily influenced by and agree with the Dutch data set as expected due to a larger number of patient encounters it contains. Where significant associations appear in the Dutch dataset without a comparable association in the smaller Malta dataset, this was reflected in some associations losing significance in the combined dataset (for example “vomiting” in the case of pyelonephritis).

Limitations
This study examined associations between RfEs and episode titles at the beginning of a new EoC for that problem. It is quite possible that the diagnosis may have been revised over time during another consultation forming part of the EoC due to a change in the presentation, or a change in the diagnostic opinion of the FD, or consequent to the results of further testing, or through an opinion expressed by another health care provider, or otherwise.

A larger dataset would have quite likely picked up more significant associations, and provided more precise estimates of effects. The observed differences in diagnostic associations between populations may thus be due more to the lack of power to define the LRs more precisely, rather than due to any real difference in diagnostic processing of such RfEs.

We hereby publish the LRs used to study and describe these diagnostic associations in two different populations along with a combined dataset, and we offer our interpretation of the strength and reliability of such diagnostic associations, summarising the empirical data in text form. We understand that others may interpret these data differently, or may choose to accept different limits for the clinical and statistical significance of such associations.

Strengths
This is a study of diagnostic associations for two common diagnoses in practice populations in very different health care settings, which has the advantage of empirical data collection and the validation of observations between two independent datasets. We analysed data on all RfEs presented and all diagnoses made in EoCs, which allows one to study any possible diagnostic association and define those which reach clinical and statistical significance. The presented data are but two examples. We also applied tight clinical and statistical significance limits to avoid describing spurious associations. The congruency of the diagnostic associations across populations sustains our confidence in their validity.

Conclusions
The significant diagnostic cues obtained from the calculations performed on the Dutch data are consistent with the available clinical literature on LRs relating to both diagnostic conditions investigated. We conclude that it is possible to generate clinically meaningful diagnostic evidence from electronic sources of patient data.

Further research in this area is important to sustain the development of FM as a clinical and academic discipline, and to inform decision support tools and systems developed for family practice. The assumptions we have made on the clinical and statistical significance limits for a diagnostic association, and the method we have used to interpret and summarise such diagnostic associations in different populations, are presented to the scientific community for discussion.

Availability of supporting data
All supporting data is provided as an additional file showing the raw excel output from the KNIME data mining process upon which Additional files 2 and 3 were prepared.
Additional files

Additional file 1: Raw exported data from KNIME data mining tool used to create Table S1 and Table S2.

Additional file 2: Table S3. Showing positive likelihood ratios for associated RfEs (label and ICPC code listed) and the episode title “UTI” in two populations.

Additional file 3: Table S4. Showing positive likelihood ratios for associated RfEs (label and ICPC code listed) and the episode title “pyelonephritis” in two populations.

Abbreviations
DDSS: Diagnostic Decision Support; EMR: Electronic Medical Record; EoC: Episode of Care; ICPC: International Classification of Primary Care; KNIME: Konstanz Information Miner; RfE: Reason for Encounter.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
JKS developed the research methodology, collected data (from Malta), analysed data, developed the research idea, and wrote the manuscript; DC developed the research methodology, analysed data and wrote the manuscript; PK, TK and RD has designed the analysis and analysed data; MK has developed the software tools for analysis. All authors read and approved the final manuscript.

Acknowledgements
This study is based on data analysis for the TRANSFoRm project. The TRANSFoRm project is partially funded by the European Commission – DG INFSO (FP7 247787).

This grant directly supported the work of authors DC, JKS, PK, TK, RD, MK and INFSO (FP7 247787).

This study was partially funded by the Health Research Board (HRB) of Ireland and the Mediterranean Institute of Primary Care, Wroclaw University of Technology, Imperial College London as project partners. Please see www.transformproject.eu for more details.

This work was partially funded by the Health Research Board (HRB) of Ireland and take full advantage of:
• Inclusion in PubMed, CAS, Scopus and Google Scholar
• Immediate publication on acceptance
• No space constraints or color figure charges
• Convenient online submission
• Thorough peer review
• Research which is freely available for redistribution

Submit your next manuscript to BioMed Central and take full advantage of:

Submit your manuscript at www.biomedcentral.com/submit

Submit your manuscript to BioMed Central

Submit your manuscript to BioMed Central

Author details
1Mediterranean Institute of Primary Care 19, Triq ir-Rand, Attard, Malta. 
2Department of General Practice, HRB Centre for Primary Care Research, Beaux Lane House, Lower Mercer Street, Dublin, Ireland. 
3Wroclaw University of Technology, Wybreeze Wyspianskiego 27, 50-370 Wroclaw, Poland. 
4Imperial College London, London, UK. 
5Wolfson Chair of General Practice, King’s College London, Capital House, Guy’s Hospital, London, England.

Received: 27 November 2014 Accepted: 27 April 2015 Published online: 16 May 2015

References

7. TRANSFoRm Project. [http://www.transformproject.eu/]
8. TRANSFoRm Diagnostic Associations Rules Viewer. [http://156.17.131.239/RulesAssessment]
13. KNIME Homepage. [http://www.knime.org/]
Translational Medicine and Patient Safety in Europe: TRANSFoRm—Architecture for the Learning Health System in Europe

Brendan C. Delaney,1 Vasa Curcin,1 Anna Andreasson,2 Theodoros N. Arvanitis,3 Hilde Bastiaens,4 Derek Corrigan,5 Jean-Francois Ethier,6 Olga Kostopoulou,1 Wolfgang Kuchinke,7 Mark McGilchrist,8 Paul van Royen,4 and Peter Wagner9

1King’s College London, London SE1 3QD, UK
2Karolinska Institutet, 14183 Stockholm, Sweden
3University of Warwick, Coventry CV4 7AL, UK
4University of Antwerp, 2610 Antwerp, Belgium
5Royal College of Surgeons of Ireland, Dublin 2, Ireland
6INSERM, 6 Paris, France
7University of Düsseldorf, 40225 Düsseldorf, Germany
8University of Dundee, Dundee DD2 4BF, UK
9Quintiles GmbH, 63263 Neu-isenberg, Germany

Correspondence should be addressed to Brendan C. Delaney; brendan.delaney@kcl.ac.uk

Received 16 February 2015; Accepted 8 June 2015

Academic Editor: Danielle Dupont

Copyright © 2015 Brendan C. Delaney et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Learning Health System (LHS) describes linking routine healthcare systems directly with both research translation and knowledge translation as an extension of the evidence-based medicine paradigm, taking advantage of the ubiquitous use of electronic health record (EHR) systems. TRANSFoRm is an EU FP7 project that seeks to develop an infrastructure for the LHS in European primary care. Methods. The project is based on three clinical use cases, a genotype-phenotype study in diabetes, a randomised controlled trial with gastroesophageal reflux disease, and a diagnostic decision support system for chest pain, abdominal pain, and shortness of breath. Results. Four models were developed (clinical research, clinical data, provenance, and diagnosis) that form the basis of the projects approach to interoperability. These models are maintained as ontologies with binding of terms to define precise data elements. CDISC ODM and SDM standards are extended using an archetype approach to enable a two-level model of individual data elements, representing both research content and clinical content. Separate configurations of the TRANSFoRm tools serve each use case. Conclusions. The project has been successful in using ontologies and archetypes to develop a highly flexible solution to the problem of heterogeneity of data sources presented by the LHS.

1. Introduction

The Learning Health System (LHS) describes an approach to improve healthcare that is solidly founded on the creation and use of knowledge; “health” as opposed to “healthcare” is sometimes used to emphasise the role of consumers as cocreators and users of health knowledge [1]. The development of the LHS is a natural outcome of the evolution of evidence-based medicine (EBM). Based on the greater utilisation of electronic health records (EHRs) and on novel computing paradigms for data analysis, the LHS provides potential solutions for the glacial slowness of both the traditional research process and the research translation into improved care [2].

EBM is focused on generating medical evidence and using it to make clinical decisions. The highest level of
evidence, level 1 evidence of the effectiveness of a health-care intervention in EBM, consists of a meta-analysis of randomised controlled trials (RCTs) [3]. However, RCTs are complex and extremely expensive, the result being that much of healthcare remains unsupported by high quality evidence. Furthermore, RCTs themselves are prone to bias and manipulation in the choice of eligible subjects, comparators, and outcome measures [4]. One solution has been to carry out light touch and simple, termed “pragmatic” RCTs with very inclusive eligibility criteria and followup via routine data collection. It is those kinds of RCTs that lend themselves most to incorporation into a LHS.

There is also potential to replace RCTs with analysis of routine data, using techniques such as instrumental variables and propensity scores to control for bias [5]. Much future research is needed to define when routine data could be a sufficient answer to a problem and when an RCT is required. Furthermore, healthcare practice is not solely limited to interventions, but diagnosis and prognostication play essential parts and are underpinned by prospective cohort evidence. Again, routine data could play a significant role in replacing time-consuming and costly cohort designs.

Primary healthcare is the first point of contact with health services of patients with undifferentiated problems and also provides continuing care for patients with chronic diseases and follows families from “cradle to grave.” These functions present a particular problem for EBM. The vast majority of research, be it diagnostic or intervention based, takes place in specialist centres and in highly selected populations [6]. Diagnostic features are not portable across populations with different prevalence and spectrum of disease. Likewise, patients in RCTs are younger and fitter, take fewer drugs concurrently, and have less comorbidity than typical primary care populations. Therefore, many RCTs suffer from limited external validity [7].

Even if appropriate research evidence exists, it is unlikely to be available at the point of care. Early formulations of EBM typically applied to the highly motivated clinician who formulates questions during clinical practice and searches for evidence. Indeed, Professor Sackett’s team at Oxford developed an “evidence cart” for ward rounds, with a copy for MEDLINE and a projector to assist in this process in real time [8]. Over the subsequent years, the process of knowledge translation has become formalised: guidelines are explicitly built on systematic reviews of the best available evidence and are refined down to a series of statements to support clinical care, with an associated level of supporting evidence and strength of recommendation [9]. However, even in countries like the UK, where a national agency (National Institute for Health and Care Excellence) is funded to carry out this process, guidelines may only be updated once in a decade. Increasingly, the number of potential guidelines applicable to a problem of memory and prioritisation for the clinician, let alone the patient. The LHS offers a potential means of using highly advanced electronic triggers to help with advising when one treatment or diagnosis is favoured. It should also be possible to reintroduce patient choice by explicit weighting of options using patient-derived outcome data.

The LHS concept is still in its infancy, and much needs to be done to explore and demonstrate the potential for using an advanced digital infrastructure to support the LHS. The FP7 TRANSFoRm project (http://www.transformproject.eu/) was funded via the Patient Safety Stream of ICT for Health. Efficient research design and knowledge translation are a core underpinning of safe clinical practice. It is not good enough to simply avoid error, defined as care that falls well below the average standard, but clinicians should be seeking optimal care for their patients. The LHS, at its barest essential, is all about promoting optimal care. The TRANSFoRm project aimed to develop and demonstrate methods, models, standards, and a digital infrastructure for three specific components of the LHS:

1. genotype-phenotype epidemiological studies using multiple existing primary care and “biobank” genomic datasets;
2. RCTs with both data and trial processes embedded within the functionality of EHRs and the ability to collect Patient Reported Outcome Measures (PROMs) on demand;
3. decision support for diagnosis, based on clinical prediction rules (best diagnostic evidence) and fully integrated with a demonstrator EHR system.

2. Methods

Each specific clinical “use case” (shown below) served four purposes: initial requirements elicitation; detailed modelling of infrastructure and required data elements; design of concurrent validation and evaluation studies; and final clinical demonstrations. 21 partner organisations in ten EU member states took part in the project, over five years. At the time of writing, the project has 11 months to run and the final evaluation and clinical studies are about to commence.

TRANSFoRm Use Cases

Diabetes Use Case. The aim of the Diabetes use case is to enable a distributed query to look for eligible patients and extract data from multiple federated databases. In the pilot study, the query will define patients and data to support analysis of the relationship between well-selected single nucleotide polymorphisms (SNPs) in type 2 diabetic patients and the response to sulfonylurea.

GORD Use Case. The aim of the GORD use case is to investigate the effectiveness of on-demand versus continuous use of proton pump inhibitors on reflux symptoms, quality of life, and self-rated health in patients with gastroesophageal reflux disease in primary care. The study will be conducted in five localities (UK: two vendors, Poland, Netherlands, and Crete) and it will aim to recruit, randomise, and follow 700 patients at 40 primary care centres using the clinical trial application.

Diagnosis Use Case. The aim of the diagnosis use case is to provide integrated point-of-care decision support for patients
presenting with chest pain, abdominal pain, and shortness of breath.

TRANSFoRm aims to produce a highly flexible infrastructure that presents the lowest possible barriers to entry for EHR systems and datasets, but at the same time it makes the maximum use of the existing data standards and methods for managing heterogeneity, both structural and terminological, between data sources. A basic principle of the TRANSFoRm project was to use available standards and models as much as possible and integrate them into the TRANSFoRm infrastructure. It was decided early on in the project that TRANSFoRm would take a model-based approach, using 4 models to capture (1) clinical meaning, (2) research meaning, (3) provenance, and (4) diagnostic meaning. The latter is essentially a subset of the clinical model, but it was modelled separately for efficiency. The archetype approach of constraining one model against the other, in a two-level design (clinical and research), was used to describe data elements [10]. Where available, existing tools for building and maintaining models as an ontology were used, although we presented a novel use of LexEVS, which we employed to support both structural and semantic models [11].

Clinical concepts were modelled using an ontology (termed the Clinical Data Information Model, CDIM) [12]. Additional semantic detail for data elements was expressed by using LexEVS to support binding of terminology terms to CDIM expressions. For representation of research processes, we extended an existing domain model, the Primary Care Research Object Model, adding objects primarily in the clinical area [13]. The resulting Clinical Research Information Model (CRIM), in conjunction with CDIM, enabled a two-level archetype to be defined for each required data element in the use cases. In order to define case report forms and study designs for the RCT, we used the CDISC ODM and SDM standards, but adding an archetype approach for the description of the data element “payload” [14].

The intention from the outset with TRANSFoRm was that all models would be published, standards would be reused and adapted as required, the software would reuse the existing open source components, if available, and all TRANSFoRm software components would be made available as open source tools under an “Apache” license. We believe that the value lies in the data and the knowledge generated from it and that amortizing the infrastructure can only act as a potential barrier for realising the value of the data/knowledge.

Evaluation of TRANSFoRm will consist of a technical validation of the TRANSFoRm tools and three clinical and sociotechnical evaluation studies. For the DSS, an evaluation of the system, integrated with the In Practice Systems Vision 3 EHR system, is underway. General practitioners are conducting a simulated clinical session with actors simulating patients presenting with carefully prepared test problems. This is a within-subjects design, with the cases solved first without and then with the DSS and the primary outcome being accuracy. We also measure usability and amount of information coded into the EHR. The Diabetes use case is being evaluated on the basis of performance, as judged by users, of the system in selecting and extracting data from five databases. Accuracy of selecting eligible patients by users employing the TRANSFoRm Query Workbench will be measured. The GORD (gastroesophageal reflux disease, a disorder caused by the retrograde flow of gastric contents from the stomach into the esophagus, causing symptoms and/or mucosal damage) study is being conducted as a full clinical RCT (individual subjects randomised) with a nested evaluation study. Principal outcomes of the clinical study are symptom profiles and quality of life measured by PROMs (Patient Reported Outcome Measures) collected on smartphones via a dedicated TRANSFoRm mobile data collection app. The sociotechnical evaluation is a nested cluster trial and will compare recruitment rates, completeness of data, and costs of the TRANSFoRm system compared to usual practice, in this case, a simple web form for the clinical measures and paper questionnaires for the PROMs. The results of the three TRANSFoRm evaluation studies will be available in late 2015.

3. Results

The TRANSFoRm software ecosystem is comprised of a set of generic middleware components that provide essential shared functions for the LHS applications built in TRANSFoRm, namely, secure data transport, authentication, semantic mediation, and data provenance (with respect to processing of data within TRANSFoRm). As LHS is characterized by routine production, transformation, and dissemination of data and knowledge, secure channels and reliable authentication are necessary to ensure confidence and buy-in by the data owners. The data itself resides in a vast array of distributed repositories that vary both in structure and in terminology, making data interoperability a key requirement that TRANSFoRm delivers using a semantic mediation approach combined with the standard data connectivity module (data node connector: DNC). The DNC implements data interoperability, as well as managing workflow processes and data extraction for participating EHRs and data sources, as discussed in the next section. Different flavours of DNC operate in epidemiology and RCT use cases, as the RCT DNC has to support additional requirements of the RCT workflow. Data provenance capture in TRANSFoRm implements traceability, which is necessary both to support trust and transparency and to enable learning and improvement in LHS processes.

On top of these shared components, three application specific tools were built to support the use cases: epidemiological study query workbench, clinical trial monitoring tool, and a diagnostic support plugin for EHR systems.

The high-level overview of the software components is shown in Figure 1.

4. Epidemiological Study Application

The epidemiological study TRANSFoRm software configuration (Figure 2) is used in the genotypic-phenotypic T2D study use case and consists of tools for secure, provenance-enabled design and execution of eligibility queries and data extractions from heterogeneous data sources. Eligibility queries are
FIGURE 1: High-level software components.

FIGURE 2: Epidemiological study configuration annotated with steps in the query process.
formulated by the researcher in the query workbench (QWB) web tool (Figure 3) using model-based constructs (Figure 2, step 1). QWB users enter clinical terms into the system which then presents the user with a list of corresponding concepts from standard terminologies and classifications (Figure 4). The researchers are able to use a data quality tool, storing metadata about available practices and data that reside in them, to restrict the search to practices with a high registration percentage of the variables targeted in the study (step 2). The queries are dispatched to the data sources via the middleware (step 3) to the local data node connector. This is a TRANSFoRm component that sits at the data source and translates the generic CDIM-based query into a local representation using the semantic mediator component (step 4) and subsequently presents that locally interpretable query either to the data source directly or to a human agent for final approval (step 5), before returning the result. Three types of queries are supported: patient counts, flagging patients, and data extraction. Results of count and flag queries are sent back to the query workbench via the middleware (step 6a) and can be viewed by the researcher in the QWB web tool. The patient data extraction result is passed to a safe haven (step 6b), accessible only to the authorised researcher, using the appropriate secure data transport mechanism.

5. Clinical Trial Application

The clinical trial software configuration (Figure 5) is used in the GORD use case and consists of components needed for design, deployment, and collection of trial data, backed by provenance and secure authentication framework for researchers. The trial data collection is supported using electronic Case Report Forms (eCRFs) and Patient Reported Outcome Measures (PROMs). The former are filled in via a web browser by the clinician, while the latter are completed by the patients using either web or mobile devices. Also supported is the orchestration of data collection across multiple clinical sites where the trials are taking place.

The TRANSFoRm architecture delivers important components of clinical trials: patient eligibility checks and enrolment, prepopulation of eCRF data from EHRs, PROM data collection from patients, and storing of a copy of study data in the EHR. The key component of the architecture is the TRANSFoRm Study System (TSS) that coordinates study events and data collections, using HTML form templates with bound queries for preloading data from the EHR. The studies, represented using a custom extension of CDISC SDM/ODM standard, are loaded into the TRANSFoRm Study System (step 1). Whenever an interaction is required between the Study System and EHR, for example, eligibility checks or partial filling of eCRF forms form EHR data, a query is fired off to the EHR via the data node connector (step 2). As in the epidemiological study configuration, the DNC acts as a single point of contact of TRANSFoRm components and the local EHR. In addition to translating and sending queries to the EHR (step 3), the DNC acts as a web server that displays eCRF forms for the clinician to fill with study-required information. The form is submitted to both the study database and the EHR for storage considering requirements for eSource data use in clinical trials (step 4). The message protocol for this interaction is currently undergoing comparison evaluation with the IHE standards [17]. The PROM data is collected directly from the patients using web or mobile devices (step 5). The software configuration for the GORD study undergoes a formal Computer System Validation (CSV) process including qualifications for installation, operation, and performance to ensure that study system and study process have been Good Clinical Practice- (GCP-) validated prior to being employed in the GORD clinical trial use case. Because of the narrow connection between EHR and study system, part of GCP-validation is the assurance of data privacy and confidentiality of the personal patient data.

6. Diagnostic Support Application

Diagnostic support software configuration (Figure 6: diagnostic support configuration) consists of tools for mining new rules from health data sources and managing their deployment into the knowledge base, upon which an evidence service is operating to drive a diagnostic support tool embedded into a local EHR system.

The primary function of the tool is to suggest to clinicians diagnoses to consider at the start of the clinical encounter based only on the existing information in the patient record and the current reason for encounter [18]. It also allows bottom-up input of observed patient cues (symptoms and signs), independent of associated diagnosis, or top-down
drilling into and selection of cues supporting specific diagnoses.

The rules used in the diagnostic process are generated by data mining tasks (step 0), which get manually curated and fed through the evidence service into the Clinical Evidence Repository. When the patient presents, the cues entered or selected are then used to dynamically rank the potential differential diagnoses (Figure 7). This is done by the DSS plugin embedded into the EHR, sending data to the evidence service (step 1), which queries the rules stored in...
### Table 1: A table of outputs and exploitation plans.

<table>
<thead>
<tr>
<th>TRANSFoRm output</th>
<th>Exploitation plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Privacy model: a “zone” model with an explicit method of graphically depicting the zones and operation of filters between zones</td>
<td>Published method [15]</td>
</tr>
<tr>
<td>(2) Provenance infrastructure: based on the Open Provenance Model [REF], each infrastructure component captures a provenance trace that enables reconstruction of an audit trail for any given data element</td>
<td>Published method [16]</td>
</tr>
<tr>
<td>(3) Clinical prediction rule ontology based web service</td>
<td>The diagnostic ontology has been made available as a public download in OWL format on the TRANSFoRm website (<a href="http://www.transformproject.eu/">http://www.transformproject.eu/</a>). A future project is required to extend the data beyond the three initial reasons for encounter</td>
</tr>
<tr>
<td>(4) Research data model</td>
<td>CDIM [12] and CRIM [13] have been published. A full description of the use of CDIM and CRIM in the construction of data node connectors will be published and made available on the TRANSFoRm website</td>
</tr>
<tr>
<td>(5) eCRF</td>
<td>Extension of CDISC ODM and SDM by the incorporation of archetypes with references to the CRIM and CDIM models will be published and discussions are ongoing with CDISC regarding future incorporation into the standards. A reference implementation of the clinical trial system will be maintained within the European Institute. At present, individual archetypes have to be written by hand; discussions are in hand for the production of an archetype authoring tool</td>
</tr>
<tr>
<td>(6) Data federation</td>
<td>A reference implementation of the genotype-phenotype study system will be maintained within the European Institute. Search authoring tools will be available open source</td>
</tr>
<tr>
<td>(7) DSS integration</td>
<td>The DSS is currently integrated with the InPS Vision 3 system. Further work is required to move this to a data node connector/CDIM-based flexible system</td>
</tr>
</tbody>
</table>

### 7. Conclusions

TRANSFoRm demonstrated how a Learning Health System can be implemented in European clinical research and practice. The full list of project outputs and the exploitation plan for each are shown in Table 1 and promoted via an open source model. TRANSFoRm will be a full participant in the European Institute for Innovation through Health Data and will make its tools and models available via the institute. In addition, we are internationally active as participants and promoters of the Learning Healthcare System. Via the LHS, we are publishing models, standards, and tools to the world research community. The UK serves as an exemplar of our business model, with multiple EHRs participating in the project as well as the Medicines and Healthcare Products Regulatory Agency, Clinical Practice Research Datalink (CPRD). CPRD currently extracts data from practices to a total population of 8 million and links them to 20 other health datasets. CPRD will be using the TRANSFoRm clinical trial tools, in conjunction with additional reworking by a commercial
software vendor to create a full EHR-embedded clinical trial facility for the UK Clinical Research Network.

**Disclaimer**

The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

**Conflict of Interests**

The authors declare that there is no conflict of interests regarding the publication of this paper.

**Acknowledgments**

This project has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under Grant Agreement no. 247776. Additional support was received by the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy’s and St Thomas’ NHS Foundation Trust and King’s College London.

**References**


Classification Method for Differential Diagnosis Based on the Course of Episode of Care

Adrian Popiel¹, Tomasz Kajdanowicz¹, Przemyslaw Kazienko¹, Jean Karl Soler², Derek Corrigan³, Vasa Curcin⁴, Roxana Danger Mercaderes⁴, and Brendan Delaney⁵

¹ Institute of Informatics, Wroclaw University of Technology, Wroclaw, Poland
tomasz.kajdanowicz@pwr.wroc.pl
² Mediterranean Institute of Primary Care, Attard, Malta
³ Department of General Practice, Royal College of Surgeons in Ireland, Ireland
⁴ Department of Computing, Imperial College London, United Kingdom
⁵ Kings College London, United Kingdom

Abstract. The main goal of the paper is to propose a classification method for differential diagnosis in primary care domain. Commonly, the final diagnosis for the episode of care is related with the initial reason for encounter (RfE). However, many distinct diagnoses can follow from a single RfE and they need to be distinguished. The new method exploits the data about whole episodes of care quantified by individual patients’ encounters and it extracts episode features from electronic health record to learn the classifier. The experimental studies carried out on two primary care dataset from Malta and the Netherlands for three distinct diagnostic groups revealed the validity of the proposed approach.

Keywords: Classification, Differential Diagnosis Classification, Episode of Care Diagnosis.

1 Introduction

Modern-day human is surrounded by a lot of information. However in short time he learnt how to: use it, put in data warehouses, explore, mine and make profit on it. As far as we deal with these techniques on everyday basis in management and marketing, they are rarely used in primary care.

It is addressed in the paper an interesting problem of differential diagnosis classification that can support the decisions of General Practitioners (GPs). This problem concerns the diagnosis for diseases, which have common reason for the first encounter but end with different, in some cases very serious diagnosis.

This paper provides a proposal of classification method for differential diagnosis in three groups of diseases, for which decision support has a great value for a General Practitioner. The method takes into account medical history encapsulated within the episode of care and results with final episode diagnosis. In order to infer the diagnosis the method uses direct attributes that describes patient(age group, referrals to primary care and specialists) as well as attributes derived from patient episode of care history(for instance number of distinct diagnosis and average number of reasons for encounters of various types).
Reason for encounter (RfE) is a factor that starts encounter with GP. It could be in a form of symptom or complaint, also request for an intervention for example prescription, advice or referral, in some cases even diagnosis from previous encounter can be a RfE. However, it needs to be underlined that in interpretation RfE differs from diagnosis. In particular RfE starts an encounter and diagnosis is a final result of a medical treatment in this encounter[1].

The paper provides concise presentation of related work in the field of medical decision support in Section 2 and description of decision support problem for General Practitioners in Section 3. It is proposed a classification method for differential diagnosis accompanied with short presentation of used algorithms for Differential Diagnosis as well as methods of data evaluation in Section 4. Then experimental results and comparison of the methods are gathered in Section 5 and concluded in Section 6.

2 Related Work

The general source of information for medical decision making is an electronic health record (EHR). In the recent years in the United States it was invested nearly 50 billion dollars to create EHRs as such data repositories are expected to improve quality care and reduce costs. One of the main goals of EHRs is to improve the quality of diagnosis and overall treatment providing EHR based information technology tools, especially decision support systems. Differential diagnosis in primary care might be complex and in some cases requires long diagnostic process. Providing decision support systems might help General Practitioners feel more comfortable and supported in quick diagnostics. Yet practitioners might appreciate additional knowledge source especially when they must make critical diagnosis. Nevertheless, decision support systems have today several number of limitations, but implemented with practical and evidence based approach, can be an important enhancement in primary care decision making process [2].

However, there exist some opposite conclusions like these stated by Romano and Stafford who provided conclusions that EHR is not associated with better quality in health centres, so that make concerns about the ability of health information technology to provide better quality in primary care [3].

Partial answer for that problem brings Kortteisto, who described process of implementation of decision support and medical data recording system in health center in Finland [4]. It was highlighted that such systems must be adopted well by employees of health centres in order to bring additional value. Even thought the decision support system was introduced to GPs and made familiar for them, practitioners recorded reasons for encounters (RfEs) or diagnosis when patient has already left. This caused wrong qualification of RfEs and errors in data filling. Moreover, in the proposed system it was in fact to much support from the system and it was hard to make profit from such amount of knowledge.

Important part of medical decision support systems is technology, methods and algorithms behind them. Yoo et al. bring great analysis of data mining techniques used
in health care [5]. In their work there were presented data mining techniques and algorithms and as well as contexts of their usage in biomedical and healthcare studies.

Li et al. focused on efficient discovery of risk patterns in medical data [6]. Their algorithm quickly and efficiently discovers cohorts of patients that are vulnerable to a risk outcome. They also compared their method with decision trees and association rules that showed that discovered by their method risk patterns had much more quality than patterns brought by classic data mining methods.

Yang and Wang used Random Forest to design new classifier that create more reliable classes, because in medicine risk of misjudgment is very costly. Their classifier take into account cost of misjudgment and predefined confidence level for each class [7].

Summarizing, the problem of differential diagnosis was rarely touched on in the context of information technology solutions. Especially nobody bind this topic with course of episode of care.

3 Problem Description

In many electronic health record systems (EHRs) gathering information for primary health care, the data records are composed of three main parts: (1) demographic data about the patient (age, sex, region, population), (2) provenance of data (lineage of the processes in a EHR system) and (3) series of following patient encounters. Every such a sequence starts at the first visit when the patient provides his or her primary (initial) reason for encounter (RfE) and terminates with the last encounter containing the final diagnosis, Fig. 1.

From the reasoning point of view the most important episode features are: (1) initial reason for encounter (RfE) expressed by the patient at the first encounter, (2) diagnoses, symptoms and procedures registered during all intermediate visits and (3) the final diagnoses fixed at the last encounter – it can already be fixed during the previous encounters. Both RfEs and diagnoses in primary health care can be coded by means of ICPC2 - International Classification of Primary Care, 2nd Edition developed by WONCA International Classification Committee (WICC) [8], available at http://icpc.who-fic.nl/browser.aspx. The main difference between RfE and a diagnosis is their creator: reasons for encounter are provided by patients themselves, whereas diagnoses are made the general practitioners (GPs). The initial RfE remains static for the whole episode but the diagnosis may change even at each encounter according to new physician findings.

Based on the ICPC2 coding schema some aggregated episode of care features can be derived from the component encounters, Fig. 2, see Sec. 4.2 for details.

The main goal of differential diagnosis is to distinguish final diagnoses that can be be initiated by a given reason for encounter. In other words, the main goal of analysis of data sets collected in electronic health records for primary care is discover general rules that would help family doctors to be aware of various results (final diagnoses) at the time when a patient tells their complains (RfE), Fig. 2. Hence, differential diagnoses distinguish various final diagnoses that can be caused by single reason for encounter.
Fig. 1. The data-driven organisation of an episode in primary health care

Fig. 2. Differential diagnosis for a single reason for encounter (RfE)
4 Classification Method for Differential Diagnosis

The proposed method for differential diagnosis classification is based on the data extracted from the electronic health record database (EHR). It contains records on individual patient encounters that can be aggregated into single episode of care records (EoC), see Fig. 1 and 2.

The entire classification method for differential diagnosis consists of five major steps, Fig. 3. These are: (1) identification of whole episodes of care (EoC) from the series of individual encounters taken from source EHR (in experimental studies TransHIS data was used as the source data set); (2) extraction of features describing EoC, including primary/initial reason for encounter (RfE) that starts episode of care (in order to select subjects that are taken into account), final diagnosis, i.e. diagnosis from the last encounter used for the classification output; (3) feature selection, (4) learning of the classifier (building the model) and (5) validation (at research) or testing (in decision support system). The method utilizes the original medical data of encounters arranged in episodes of care. In general, the encounter data contains diagnosis (Dia) that was assigned at the encounter by GP alongside with reasons for encounter (RfE) told by the patient, symptoms (Anam) and procedures (Proc) that were prescribed before a given encounter. The information about the encounter is followed by demographics of patient (age group, sex, etc.) as well as referrals to specialists to be undertaken after the encounter. Multiple encounters constitute an episode of care. The episode of care is quantified with final diagnosis called episode diagnosis (EpisodeDia). Moreover, depending on the type of episodes they can have distinct status: new or pre-existing, where an old problem is presented to GP. The method considers only new episodes of care. Moreover, the data describing EoC should have common quantification for RfE, symptoms, diagnosis and procedures, e.g. ICPC2 (International Classification of Primary Care 2nd Edition [8]). The following steps are described below more in-depth.

![Fig. 3. The general schema of classification method for differential diagnosis](image-url)
4.1 Initial RfE, Diagnostic Groups

Based on raw data (consisting of encounters arranged in episodes of care) only new episodes of care were selected (step 1 in Fig. 3). Fortunately, TransHIS data set provides a indicator of episode status. In other case this would require additional episode status discovery. Since the main purpose of the method is to distinguish diagnosis for some selected primary reason for encounter, the episodes were partitioned in three groups, based on their initial RfE (according to ICPC2 codes). These are RfE known as relatively hard to make a correct diagnosis at the beginning (it means that they especially require differential diagnosis):

**Group 1**, initial RfE: D01 - Abdominal Pain, with final episode diagnosis:

- D73 - Gastroenteritis presumed infection
- D88 - Appendicitis
- D93 - Irritable Bowel Syndrome
- D99 - Disease digestive system, Chrons disease
- U70 - Pyelonephritis/pyelitis
- U71 - Cystitis/urinary infection other
- W80 - Ectopic Pregnancy
- X74 - Pelvic Inflammatory Disease
- X77 - Malignant neoplasm genital other
- X81 - Genital neoplasm other/unspecified, Ovarian Cancer

**Group 2**, initial RfE: A11 - Chest Pain, with final episode diagnosis:

- A70 - Tuberculosis
- K74 - Ischaemic heart disease
- K93 - Pulmonary Embolism
- R81 - Pneumonia
- R84 - Malignant neoplasm bronchus/lung
- R85 - Malignant neoplasm respiratory, Mesothelioma
- R88 - injury respiratory other, Pneumothorax
- R96 - Asthma

**Group 3**, initial RfE: R02 - Shortness of breath/dyspnoea, with final episode diagnosis:

- K77 - Heart Failure, Right Ventricular Failure
- K82 - Pulmonary heart disease, Cor Pulmonale
- K83 - Heart valve disease, Aortic Stenosis
- R78 - Acute Bronchitis/bronchiolitis
- R95 - Chronic obstructive pulmonary disease

Experimental studies were restricted only to above three groups.
4.2 Episode of Care Aggregated Features

Commonly, the basic form of raw data provides only basic information about episodes such as episode diagnosis, certainty of this diagnosis and initial status. In order to derive more episode specific features, enumeration of encounters within the episode was proposed. Then, some aggregated features describing the whole episode were computed (step 2 on Fig. [3]). The following were sued for that purpose: the number of specific encounters, their percentage, average, variance, standard deviation, maximum, minimum and median of RfE, symptom and diagnosis within a particular ICPC2 diagnosis group. For instance, the above mentioned aggregations of RfE were calculated for all RfE in the episode with type D (Digestive, ICPC code starting with 'D'), another set of features for type A, B, H, K, etc. In general, this feature derivation will result in 17 (the number of distinct diagnosis groups in ICPC2) attributes for RfE, symptom and diagnosis for each of aggregation method. It provides all together 408 attributes describing each episode.

Moreover, a similar feature derivation can be applied to medical procedures, additionally enumerated in five groups based on procedure number:

1. 30-49 - Diagnostic and preventive procedures
2. 50-59 - Treatment procedures, medication
3. 60-61 - Test results
4. 62 - Administrative
5. 63-69 - Referral and other reasons for encounter

This will result in 595 new attributes. Next, it can be calculated 5 additional attributes about the course of episode of care, in particular the number of RfE’s, symptoms, diagnosis and procedures in the episode. In order to quantify the frequency of encounter during the episode there are extracted number, average, variance, standard deviation, maximum, minimum and median of days, weeks and years between encounters in the episode – all together 21 attributes.

Eventually, due to the fact that number of attributes is large, feature selection method is applied (step 3 in Fig. [3]) [9]. It happens often that there are rejected maximum, minimum and median values as they are not presenting good discrimination and such a selection was performed within experiments.

5 Experimental Study

5.1 Experimental Scenarios

The main purpose of performed experiments was to evaluate the predictive accuracy of the method for differential diagnosis. Moreover it was expected to observe whether results of the method can be used by general practitioners. For the experiments, the tool KNIME[10] version 2.7.3 with Weka[11] 3.6 was used. As classifiers the following models were utilized: J48 (C4.5), Random Forest (RF) and Naive Bayes (NB). In order to learn and validate these classifiers (they were learnt with default WEKA parameters), 10-fold cross-validation procedure was applied(steps 4 and 5 in Fig. [3]). For evaluation purpose, the F1-score (also named as F-measure or F-score) was used. It includes both
precision and recall to generate the score for the model. To define precision and recall in our case, there is a need to provide such values as: true positives (tp), false positives (fp) and false negatives (fn). Terms positives and negatives are related to classifier results, whereas true and false values are connected with prediction verified by some kind of external observation. So we can describe true positive as correct result in both ways, true negative as expected misjudgement (or lack of result), false positive represents correct but unexpected prediction and false negative as straightforward missing result.

Hence, precision is defined as follows [9]:

$$\text{precision} = \frac{tp}{tp + fp},$$  \hspace{1cm} (1)

recall as [9]:

$$\text{recall} = \frac{tp}{tp + fn},$$  \hspace{1cm} (2)

and F1-score as:

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$  \hspace{1cm} (3)

5.2 Datasets

The experiments were performed on TransHIS data set collected in the Netherlands[1]. There were extracted three groups of episodes of care as described in Section 4.2. Group 1 consisted of 28 thousand of episodes, Group 2 and Group 3 contained 39 thousand of episodes. The data sets were processed with 10-fold cross validation.

5.3 Results

The experimental result are presented in Figures 4a, 4b and 4c. Vertical axis represents F1-score. Horizontal axis presents diagnosis from proper group. The gray mark on all figures represents the baseline level of F1-score. This is a random baseline above which the differential diagnosis outperform random classification.

In almost all cases Naive Bayes classifier (NB) returns much less satisfying results, but that was expected since it is the simplest algorithm with two very optimistic assumptions. C4.5 and Random Forest provide quite similar results, even though Random Forest is slightly better. However, C4.5 has a great advantage: it provides human understandable rules that can be interpreted by physicians. That is why C4.5 results are treated as the main achievement of the research. It can be directly used in primary health care as help for general practitioners in their diagnosis.

For group 1, Fig(a) all algorithms distinguished all diagnosis. All of the classifiers resulted with higher F1-score than a baseline. The same situation exists in group 2(b) and group 3(c). The worst results were obtained with Naive Bayes classifier.

The best results were achieved for group 3. However, for diagnosis K82, Naive Bayesian classifier provided results close to the random baseline. The proposed method always provides a good prediction, for K82 itself, but there were too few episode diagnosis with K82 (only 13 cases, so there were only 13 true positives). The algorithm
Fig. 4. Results of differential diagnosis for all three algorithms (decision tree C4.5, random forest - RF, naive Bayes - NB) for distinct diagnostic groups

also made much more predictions as K82 for other diagnosis - K77 and K88 (so, there were much more false positives: as of 620). As a result, precision (eq. 1) was low and F1-score (eq. 3) remained close to the baseline.

6 Conclusions and Future Work

A new method for differential diagnosis in primary health care was proposed in the paper. Appropriate modelling of source data (especially aggregated feature extraction) facilitated achieving good results in distinguishing diagnosis for a given reason for
patient encounter. Commonly, the best quality results were obtained for the random forest classifier even though C4.5 decision tree was not much worse.

The results can support general practitioners in their diagnosis making based only on simple and general reason for encounter like ‘abdominal pain’ or ‘chest pain’.

Future work will focus on analysis of other diagnostic groups, their medical verification as well as application of ensemble classification methods verified on other data sets.

Acknowledgement. This work was partially supported by The Polish National Center of Science the research project 2011-2014 and the European project TRANSFoRm, FP7 247787, 2011-15, www.transformproject.eu.

References