http://researchspace.auckland.ac.nz

ResearchSpace@Auckland

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author’s right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage. http://researchspace.auckland.ac.nz/feedback

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form and Deposit Licence.

Note: Masters Theses

The digital copy of a masters thesis is as submitted for examination and contains no corrections. The print copy, usually available in the University Library, may contain corrections made by hand, which have been requested by the supervisor.
ChronoMedIt – A Computational Quality Audit Framework for Better Management of Patients with Chronic Disease

by

Thusitha Dananjaya De Silva Mabotuwana

Abstract

Chronic disease is a leading cause of death worldwide accounting for around 60% of all deaths. An important aspect of successful chronic disease management is quality audit and feedback to clinicians. However, due to the complex temporal relationships inherent in chronic disease, formulating clinically relevant queries is difficult using the querying tools often built into commercial practice management systems.

The onset of this PhD research involved working with staff of a general practice clinic to develop a set of explicit quality audit indicators for blood pressure control. Eight indicators were identified as most relevant to the practice. The ability to compute these indicators reliably from routinely collected electronic medical records (EMRs) was validated by clinical panel assessment. These eight indicators informed formulation of a model of chronic disease audit with four broad classes of indicators: (1) persistence to indicated medication; (2) timely measurement recording; (3) time to achieve target; and (4) measurement contraindicating therapy.

The four broad indicator classes have been implemented within the ChronoMedIt (Chronological Medical auditIt) framework as an extensible and configurable architecture. The main components of the ChronoMedIt architecture are: an XML based specification for indicator formulation (with an associated XML-Schema), a drug and classification knowledge base maintained using Semantic Web technologies, a C# based criteria processing engine, a SQL-Server based patient database with related stored procedures and a graphical user interface to formulate queries and generate reports. ChronoMedIt can produce patient-specific audit reports as well as reports to benchmark an entire practice for a given evaluation period. A visualisation tool has been developed to provide an alternate representation of patient prescribing and measurement histories. By modifying the indicator specification and knowledge base an analyst can address a wide array of chronic disease management queries as specific instances of the four broad indicator classes. The framework’s core computation has been verified using redundant query implementations on a battery of simulated case data and is illustrated against the EMRs of several practices.
ChronoMedIt has been applied in several real-world settings; notably, identifying patients with poor antihypertensive medication adherence profiles for a feasibility study of nurse-led adherence promotion.
Acknowledgements

It has been a long journey to get to the point of this thesis submission and I would like to express my sincere thanks to several people. First and foremost I would like to express my utmost gratitude to my PhD supervisor, Prof. Jim Warren for all his support during my PhD years. I certainly cannot imagine having a better supervisor than him and I cannot thank him enough for all his guidance, patience, knowledge, motivation and integrity. Having 15+ peer-reviewed publications is by no means easy for a PhD student to achieve within three years and I attribute this success to Jim since I could not have achieved this without his constant support and feedback with quick turnaround times on manuscripts. It has been a privilege working with Jim and I would also like to thank him for giving me the fantastic opportunity to work as an Assistant Research Fellow for the National Institute for Health Innovation (NIHI).

I would also like to thank my associate supervisor A/Prof. Tim Kenealy for all his support, encouragement and guidance, especially on clinical matters.

My sincere thanks also go to our clinical collaborators, especially Dr. John Kennelly whom I have been working with from the beginning of my PhD. John has been a clinical lead and getting access to production patient data in a timely manner would have been a much more daunting task without his support. Also, special thanks to Tim and Dr. Stewart Wells for assisting me in gaining access to patient data from their respective practices.

The constant support from the NIHI team – Malcolm, Chris, Koray, Rekha, Martin, Karen, Deb and Terri is also much appreciated. Also, special thanks to Rekha whose initial work with Jim gradually led to my PhD work. My fellow PhD students – Mehnaz, Priyesh, Helen and Jose – it has been great working with you all, worrying about ‘getting our PhDs done’ as well as sharing the somewhat regular setbacks/disappointments we have had together. I am sure our PhD years will always remain a great learning experience for us all. Also, thanks to my good friends Duleepa (special thanks to you for the tips on various PhD related matters), Saeyorn, Kushil, Thiranjith, Kasun, Charaka, Ruchira and Sasanka & Sasanka for organising various get-togethers and gaming sessions; they have been a good source of distraction from continuous thesis work.
I could not have completed this research without the financial support I received from The University of Auckland Doctoral Scholarship as well as NIHI and I thank both institutes for their support. Also, I have had the privilege of attending three local conferences and three international conferences during my PhD years and this was only made possible by the kind financial backing from Jim through his HoD account as well as the conference travel support grants from BuildIT, National Heart Foundation, The Royal Society of New Zealand and The University of Auckland’s PReSS funding for postgraduate students.

Last but not least, I would like to thank my loving family – mum Karnika, dad Thusantha and sister Hasini for putting up with me over the last so many years, providing me with absolutely all the support I can think of as well as encouraging me to start a PhD when I was in two minds. Working for long hours and then chatting/playing around with my little sister when I needed a break has kept me going, and I need to thank her for that. Aththa (grand-dad) and ammamma (grandma) – your encouragement is also very much appreciated and I hope I have made you proud by continuing your line of ‘educated children’. Finally, I would not have had the energy to work with so much perseverance without the continuous love, care, patience and support of one person who always admired and appreciated what I did and that is my loving wife Venuka. Special thanks for her patience and putting up with me no matter what.

Thusitha Mabotuwana

4th May 2010
# Table of Contents

Abstract ..........................................................................................................................................................i
Acknowledgements ......................................................................................................................................iii
Table of Contents ..........................................................................................................................................v
Lists of Tables ............................................................................................................................................... ix
Lists of Figures .............................................................................................................................................. xi
Glossary of Terms....................................................................................................................................... xiv

1.0 INTRODUCTION.....................................................................................................................................1
  1.1 The Burden of Chronic Disease ........................................................................................................1
  1.2 The Opportunity ...................................................................................................................................2
  1.3 The Problem .....................................................................................................................................3
  1.4 The Research Question .....................................................................................................................4
  1.5 The Proposed Solution .....................................................................................................................4
  1.6 Thesis Outline ...................................................................................................................................5

2.0 RELATED WORK ....................................................................................................................................7
  2.1 Computer Interpretable Clinical Practice Guideline Modelling Techniques .................................7
    2.1.1 The Arden Syntax...................................................................................................................9
    2.1.2 GLIF ..................................................................................................................................... 10
    2.1.3 PROforma ........................................................................................................................... 11
    2.1.4 Asbru .................................................................................................................................. 12
    2.1.5 EON ..................................................................................................................................... 14
    2.1.6 CIG Review and Comparison Studies.................................................................................. 14
    2.1.7 LASSIE and DeGeL............................................................................................................... 17
  2.2 Guideline based Chronic Disease Management CDSSs .................................................................19
    2.2.1 ATHENA .............................................................................................................................. 19
    2.2.2 PRODIGY ............................................................................................................................. 22
    2.2.3 ASTI Project ........................................................................................................................ 25
    2.2.4 HyperCritic.......................................................................................................................... 26
    2.2.5 AsthmaCritic ....................................................................................................................... 26
  2.3 Temporal Analysis and Visualisation ..............................................................................................28
    2.3.1 Knowledge Based Temporal Abstraction and the RÉSUMÉ System................................... 29
    2.3.2 The EON System .................................................................................................................. 31
5.4 The Criteria Model........................................................................................................................105
5.5 The Computational Framework....................................................................................................108
  5.5.1 Creating Audit Criteria...........................................................................................................109
  5.5.2 The Drug and Classification Knowledge Base .................................................................111
  5.5.3 ChronoMedit’s Database Structure and Parsing BPs and Lab Test Results .................113
  5.5.4 Implementation of the Criteria Model ..................................................................................118
  5.5.5 Framework Outputs ...............................................................................................................120
    5.5.5.1 Adherence/Persistence to Indicated Medication.......................................................120
    5.5.5.2 Timely Measurement Recording ...................................................................................123
    5.5.5.3 Time to Achieve Target ...............................................................................................125
    5.5.5.4 Measurement Contraindicating Therapy ..................................................................128
  5.5.6 The Graphical User Interface ...............................................................................................130
  5.5.7 Criteria Processing with SQL or Sequential Implementation and Verification of Results 135
6.0 FRAMEWORK TESTING .....................................................................................................................137
  6.1 An Overview of Software Testing Concepts ..............................................................................137
    6.1.1 A Basic Introduction ............................................................................................................137
    6.1.2 Types of Functional Testing .............................................................................................140
      6.1.2.1 Boundary Value Analysis ..........................................................................................140
      6.1.2.2 Equivalence Class Testing .........................................................................................142
      6.1.2.3 All-Pairs Testing ........................................................................................................142
  6.2 Methods .......................................................................................................................................143
  6.3 Identification of Test Cases ..........................................................................................................143
    6.3.1 Test Cases for Persistence of Medication and Timely Measurement Recording ..........144
    6.3.2 Test Cases for Measurement Contraindicating Therapy ................................................156
    6.3.3 Test Cases for Time to Achieve Target ...............................................................................160
  6.4 Random Testing to Ensure Software Reliability ..........................................................................165
  6.5 Errors Detected ............................................................................................................................172
  6.6 Limitations of Framework Testing .............................................................................................173
7.0 APPLICATIONS USING CHRONOMEDIT .............................................................................................176
  7.1 Data Extract ................................................................................................................................176
  7.2 Identifying patients satisfying the eight audit criteria ...............................................................177
  7.3 Investigating ACEi/ARB adherence issues and the relationship to BP control in patients with hypertension and diabetes ..................................................178
  7.4 Antihypertensive adherence and impact on BP control ..............................................................180
  7.5 Interval Based Measures as Quality Indicators in Blood Pressure Management .................182
    7.5.1 Analysis Protocol ...............................................................................................................183
### Lists of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 3.1</td>
<td>Therapeutic state variables and their respective ATC codes</td>
<td>53</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Prescription durations, their therapeutic state variables and corresponding state transitions for a hypothetical patient</td>
<td>55</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>Examples of types of quality audit report statistics for patients classified with hypertension</td>
<td>58</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>The eight quality improvement criteria agreed with practice panel</td>
<td>66</td>
</tr>
<tr>
<td>Table 3.5</td>
<td>Accuracy of automated queries as assessed against final review by the practice panel</td>
<td>69</td>
</tr>
<tr>
<td>Table 3.6</td>
<td>Criteria and practice panel assessments for six-month evaluation period</td>
<td>69</td>
</tr>
<tr>
<td>Table 3.7</td>
<td>Cases where panel initially disagreed with classification by automated queries (False Positives); and final classification after review</td>
<td>70</td>
</tr>
<tr>
<td>Table 3.8</td>
<td>False Negatives</td>
<td>70</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Definitions of $\text{PDC uncapped}$ and $\text{PDC capped}$</td>
<td>103</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Different adherence calculations</td>
<td>104</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>Using elements from the Criteria Model to formulate criteria in Table 3.4</td>
<td>108</td>
</tr>
<tr>
<td>Table 6.1</td>
<td>Test cases for weak-robust equivalence class testing for medication lapse with diagnosis prior to evaluation period</td>
<td>145</td>
</tr>
<tr>
<td>Table 6.2</td>
<td>All-pairs test cases for medication lapses with diagnosis prior to evaluation period</td>
<td>146</td>
</tr>
<tr>
<td>Table 6.3</td>
<td>Test cases for weak-robust equivalence class testing for medication lapse with diagnosis during the evaluation period</td>
<td>148</td>
</tr>
<tr>
<td>Table 6.4</td>
<td>All-pairs test cases for medication lapses with diagnosis during the evaluation period</td>
<td>149</td>
</tr>
<tr>
<td>Table 6.5</td>
<td>Test cases for special value testing</td>
<td>151</td>
</tr>
<tr>
<td>Table 6.6</td>
<td>Test cases when diagnosis date is just before, equal to or just after the beginning of the evaluation period</td>
<td>151</td>
</tr>
<tr>
<td>Table 6.7</td>
<td>BVA test cases for the interaction between the lapse period and the diagnosis date or beginning of the evaluation period</td>
<td>153</td>
</tr>
<tr>
<td>Table 6.8</td>
<td>Test cases for BVA on lapse duration</td>
<td>155</td>
</tr>
<tr>
<td>Table 6.9</td>
<td>Weak-robust equivalence class test cases for the management of compelling indications</td>
<td>157</td>
</tr>
<tr>
<td>Table 6.10</td>
<td>Test cases for BVA for the management of compelling indications</td>
<td>158</td>
</tr>
<tr>
<td>Table 6.11</td>
<td>Test cases for weak-robust equivalence class testing for time to achieve target</td>
<td>162</td>
</tr>
<tr>
<td>Table 6.12</td>
<td>BVA on location of uncontrolled measurements for time to achieve target</td>
<td>163</td>
</tr>
<tr>
<td>Table 6.13</td>
<td>BVA around the minimum required number of days over which consecutively uncontrolled measurements should occur</td>
<td>164</td>
</tr>
<tr>
<td>Table 6.14</td>
<td>BVA around the minimum number of measurements required to satisfy criterion</td>
<td>164</td>
</tr>
<tr>
<td>Table 6.15</td>
<td>Operational profile for various parameters in the input space</td>
<td>168</td>
</tr>
<tr>
<td>Table 6.16</td>
<td>Details of the populated test database</td>
<td>170</td>
</tr>
<tr>
<td>Table 6.17</td>
<td>Audit criteria with parameters for random testing</td>
<td>171</td>
</tr>
<tr>
<td>Table 7.1</td>
<td>Summary of the two practice datasets for funded and enrolled patients</td>
<td>177</td>
</tr>
<tr>
<td>Table 7.2</td>
<td>Number of patients classified with hypertension satisfying the eight audit criteria</td>
<td>178</td>
</tr>
<tr>
<td>Table 7.3</td>
<td>Number of patients classified with hypertension and diabetes satisfying different adherence criteria</td>
<td>179</td>
</tr>
<tr>
<td>Table 7.4</td>
<td>ACEI/ARB adherence vs. BP in patients with hypertension and diabetes</td>
<td>180</td>
</tr>
<tr>
<td>Table 7.5</td>
<td>Antihypertensive adherence vs. BP in patients with hypertension</td>
<td>180</td>
</tr>
</tbody>
</table>
Table 7.6: PPVs of quality indicators based on past performance on same indicator............. 185
Table 7.7: PPVs between different indicators for Practice-1.................................................. 186
Table 7.8: PPVs between different indicators for Practice-2.................................................. 186
Table 7.9: Patients with hypertension satisfying different quality indicators...................... 189
Table 7.10: Patients with diabetes satisfying different quality indicators............................. 189
Table 7.11: Patients satisfying a QOF indicator (BP5/DM12), but failing any one of the three
indicators of interest............................................................................................................ 189
Table 7.12: Prescribing and dispensing frequencies for selected drugs.................................. 194
Table 7.13: Adherence levels for selected long-term medications.......................................... 194
Table 7.14: Association of individual MPR_Pr and MPR......................................................... 195
Table 7.15: Frequencies with which patients are prescribed and dispensed multiple of the
six selected drugs during the evaluation period, or with supply extending into the
evaluation period................................................................................................................. 196
Table 7.16: Prescriptions with associated first dispensing within 7 days after being prescribed...... 197
Table 7.17: Characteristics of the patients on one or more of the six long-term medications........ 198
Table 7.18: Association of prescribing based non-adherence and dispensing based
non-adherence..................................................................................................................... 202
Table 7.19: Characteristics of patients prescribed antidepressant medication......................... 203
Lists of Figures

Figure 2.1: A sample ATHENA advisory recommendation appearing as a popup window on top of the existing EMR system ................................................................. 20
Figure 2.2: A sample BP-Prescription graph produced by the ATHENA DSS ...................................................... 21
Figure 2.3: The classes of objects in the PRODIGY III Virtual Medical Record ........................................... 23
Figure 2.4: A high level view of a hypertension guideline ........................................................................... 23
Figure 2.5: The thirteen possible relationships between time intervals. X and Y denote time intervals .................................................................................................................. 28
Figure 2.6: Typical inputs to and outputs of RÉSUMÉ together with point and interval based abstractions ........................................................................................................ 31
Figure 2.7: The EON architecture ................................................................................................................ 33
Figure 2.8: The components of the IDAN architecture .............................................................................. 35
Figure 2.9: Main interface of KNAVE-II .................................................................................................. 37
Figure 2.10: The VISITORS main interface ........................................................................................... 38
Figure 3.1: Therapeutic state-transitions for antihypertensive prescribing over a 20-month period for patients with hypertension and diabetes (for transitions with ≥3 occurrences) ........... 54
Figure 3.2: An individual path diagram .................................................................................................. 55
Figure 3.3: The evaluation period specific to the quality audit report ...................................................... 57
Figure 3.4: Some of the different ways of recording prescription durations within the PMS .................. 60
Figure 3.5: Multiple classifications of a single patient. PROV indicates the GP initials ............................ 62
Figure 3.6: The data preparation process .............................................................................................. 63
Figure 3.7: Overview of the QAR generation process .............................................................................. 64
Figure 3.8: Base cases to identify a lapse in medication during the evaluation period ......................... 65
Figure 3.9: Criteria and sampling for evaluation .................................................................................. 67
Figure 3.10: The three-question assessment instrument ...................................................................... 68
Figure 4.1: Temporal cases to be taken into account when formulating a query to determine “patients classified with diabetes not on ACEi/ARB any time during the evaluation period” ................................................................. 75
Figure 4.2: Stages of data pre-processing .............................................................................................. 80
Figure 4.3: Temporal cases when determining a lapse in therapy ................................................................ 81
Figure 4.4: Creation of the next_prescription links for various drug classes ............................................. 82
Figure 4.5: The Protégé-OWL classes tab showing knowledge base concepts created to represent patient diagnoses codes and the various antihypertensive drug classes .............................................. 84
Figure 4.6: Protégé-OWL instances tab showing the populated disease management ontology .............. 85
Figure 4.7: The patient data ontology and details for a selected patient .................................................. 87
Figure 4.8: The best-practice violation taxonomy .................................................................................. 88
Figure 4.9: SQWRL query to identify hypertensive patients with a lapse in their antihypertensive therapy for over 30 days with lapse extending into or occurring during the evaluation period ........................................................................... 90
Figure 4.10: The rule/query authoring and executing environment .......................................................... 91
Figure 4.11: Visualisation of a prescription timeline for a selected patient .............................................. 93
Figure 5.1: Temporal relations to consider when calculating MPR ........................................................... 101
Figure 5.2: Prescribing patterns for a hypothetical patient ............................................................................. 104
Figure 5.3: Elements of ChronoMedIt’s Criteria Model (using UML notation) ........................................ 105
Figure 5.4: Enumerations required for different criteria elements of ChronoMedIt's Criteria Model .......................................................... 107
Figure 5.5: Components of ChronoMedIt ................................................................................................................................. 109
Figure 5.6: A snippet of the XML document showing reporting criteria to identify patients with hypertension and diabetes who have an ACEi/ARB MPR less than 80% with a lapse greater than 30 days. .............................................. 110
Figure 5.7: Two views of the drug and classification knowledge base. ................................................................. 112
Figure 5.8: Database table definitions .............................................................................................................................. 114
Figure 5.9: Sections for notes within MedTech32 ............................................................................................................. 115
Figure 5.10: A ‘blob’ representation of clinical notes ......................................................................................................... 115
Figure 5.11: Viewing lab test results within MedTech32 .............................................................................................. 116
Figure 5.12: A ‘blob’ representation of a lab test result .................................................................................................. 116
Figure 5.13: A tool to extract BPs from clinical notes and lab test results from MedTech32’s Provider Inbox .............................................................................................................. 117
Figure 5.14: Parsing lab test results ................................................................................................................................. 118
Figure 5.15: Implementation of the Criteria Model using a SQL and sequential based approach .................................................. 119
Figure 5.16: Details for a patient who satisfies criterion ‘ACEi/ARB MPR less than 80% with lapse greater than 30 days after classified with hypertension and diabetes’ .......................................................................................... 121
Figure 5.17: A prescribing chart for a selected patient ........................................................................................................ 122
Figure 5.18: A prescription chart filtered by ACEi ............................................................................................................ 123
Figure 5.19: The XML document to determine patients with hypertension who have a lapse in BP measurement for over 120 days after a high BP measurement .............................................................................. 124
Figure 5.20: Details for a patient with no BP for ≥120 days after a high BP measurement ...................................................... 124
Figure 5.21: A prescribing and a BP measurement plot .................................................................................................. 125
Figure 5.22: The XML document to determine patients with hypertension who have three or more consecutively high BP measurements over 120 days or more ................................................................................................................. 126
Figure 5.23: Details for a patient with consecutively high BP measurements .................................................................................. 127
Figure 5.24: A prescribing and a BP measurement plot for a patient with consecutively high BPs ................................................................................................................................................................. 127
Figure 5.25: The XML document to determine hypertensive patients who have a high uric acid value while on thiazide diuretics .............................................................................................................................................................................. 128
Figure 5.26: Details for a patient with high uric acid levels while on thiazide diuretics ...................................................... 129
Figure 5.27: A prescribing and a uric acid plot for a patient with high uric acid levels while on thiazide diuretics .............................................................................................................................................................................. 129
Figure 5.28: ChronoMedIt’s main user interface ...................................................................................................................... 130
Figure 5.29: The classification selector ............................................................................................................................. 131
Figure 5.30: The drug/therapy selector ............................................................................................................................ 132
Figure 5.31: Saving a criterion as an XML document ........................................................................................................... 133
Figure 5.32: Loading a previously created criterion ........................................................................................................ 133
Figure 5.33: Opening the ‘Prescription Chart’ tool ............................................................................................................... 134
Figure 5.34: Generating prescribing charts ....................................................................................................................... 134
Figure 5.35: Formulating a ‘Lapse in Observations’ criterion ............................................................................................ 135

Figure 6.1: Specified and implemented program behaviour .................................................................................................. 138
Figure 6.2: Specified, implemented and tested behaviours ..................................................................................................... 139
Figure 6.3: BVA test cases for a function of two variables ........................................................................................................ 141
Figure 6.4: Robustness test cases for a function of two variables ............................................................................................. 141
Figure 6.5: Weak-robust equivalence classes for medication lapse with diagnosis date prior to evaluation period .............................................................................................................................................................................. 144
Figure 6.6: Weak-robust equivalence classes for medication lapse with diagnosis date during the evaluation period .............................................................................................................................................................................. 147
Figure 6.7: Special cases for medication lapse when diagnosis is prior to the EP ........................................................................... 150
Figure 6.8: Special cases for medication lapse when diagnosis is during the EP ........................................................................... 150
# Glossary of Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACEi</td>
<td>Angiotensin-converting enzyme inhibitor</td>
</tr>
<tr>
<td>ARB</td>
<td>Angiotensin receptor blocker</td>
</tr>
<tr>
<td>ATC</td>
<td>Anatomical Therapeutic Chemical</td>
</tr>
<tr>
<td>ATHENA</td>
<td>Assessment and Treatment of Hypertension: Evidence-Based Automation</td>
</tr>
<tr>
<td>CDSS</td>
<td>Clinical decision support system</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Intervals</td>
</tr>
<tr>
<td>CIG</td>
<td>Clinical interpretable guideline</td>
</tr>
<tr>
<td>COPD</td>
<td>Chronic obstructive pulmonary disease</td>
</tr>
<tr>
<td>CPG</td>
<td>Clinical practice guideline</td>
</tr>
<tr>
<td>CPT</td>
<td>Current Procedural Terminology</td>
</tr>
<tr>
<td>CSV</td>
<td>Comma-separated values</td>
</tr>
<tr>
<td>CVD</td>
<td>Cardiovascular disease</td>
</tr>
<tr>
<td>CVR</td>
<td>Cardiovascular risk</td>
</tr>
<tr>
<td>DHB</td>
<td>District Health Board</td>
</tr>
<tr>
<td>eGFR</td>
<td>Estimated glomerular filtration rate</td>
</tr>
<tr>
<td>EMR</td>
<td>Electronic medical record</td>
</tr>
<tr>
<td>EP</td>
<td>Evaluation period</td>
</tr>
<tr>
<td>GLEE</td>
<td>GLIF3 Guideline Execution Engine</td>
</tr>
<tr>
<td>GLIF</td>
<td>GuideLine Interchange Format</td>
</tr>
<tr>
<td>GP</td>
<td>General practitioner</td>
</tr>
<tr>
<td>HbA1c</td>
<td>Haemoglobin A1c</td>
</tr>
<tr>
<td>HTML</td>
<td>HyperText Markup Language</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
</tr>
<tr>
<td>ICD</td>
<td>International Classification of Diseases</td>
</tr>
<tr>
<td>ICPC</td>
<td>International Classification of Primary Care</td>
</tr>
<tr>
<td>JDBC</td>
<td>Java Database Connectivity</td>
</tr>
<tr>
<td>JNC6</td>
<td>Sixth Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure</td>
</tr>
<tr>
<td>JNC7</td>
<td>Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>UTAR</td>
<td>Unexpected temporal association rule</td>
</tr>
<tr>
<td>VBA</td>
<td>Visual Basic for Applications</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

This introductory chapter presents an overview of the motivations behind this PhD research. The primary domain of clinical interest is chronic disease and therefore this chapter first gives a background related to chronic disease and the importance of its effective management. This is followed by a discussion of the need for better query formulation tools for chronic disease management – a gap the contribution of this thesis aims to narrow. A high level view of the proposed solution is then presented while a guide to the rest of this thesis is given at the end of this chapter.

1.1 The Burden of Chronic Disease

According to the World Health Organization, chronic disease was the leading cause of death worldwide in 2005 accounting for around 35 million (or 60%) of the 58 million deaths [1]. It has been projected that in 2015, 41 million (or 64%) out of all 64 million deaths will be due to chronic disease [2]. In New Zealand (NZ), chronic disease is the leading cause for hospitalisation and accounts for 70% of health expenditure and 80% of all deaths [3]. These indicate that the burden of chronic disease is a growing concern worldwide and any incremental efforts to improve the management of chronic disease will have immense population health benefits. There is a globally growing need for better management of patients with chronic disease to control the ever-increasing demands for clinical resources and also to minimise related healthcare costs. Moreover, chronic disease management introduces technical challenges associated with temporal reasoning, such as problems of computation on time
intervals and their relationships. Therefore, developing solutions that address such temporal reasoning challenges is one of the main technical focuses of this thesis.

There are many facets to improving the management of patients with chronic disease. To-date, various attempts have been made to improve the management of these patients as well as their clinical outcomes, and these include using clinical decision support systems (CDSSs), reminder systems, audit reports, increasing practitioner familiarity with clinical practice guidelines (CPGs) and providing financial incentives for following CPGs. Another important aspect is developing techniques where clinicians can actively identify patients who are (or have been) on suboptimal therapy and recall these patients to visit their general practitioners (‘GPs’ in the British, Australian and NZ systems; roughly equivalent to ‘family physicians’ in the United States [US]) who can then manage these patients accordingly. This gives rise to the concept of quality audit indicators which are specific measures used to give an indication of the quality of the patient care delivered [4]. However, formulating quality audit indicators requiring explicit treatment of temporal issues, something that lies at the very heart of chronic disease management, often poses a challenge to GPs as the process involves taking various temporal aspects into consideration, such as the interactions between prescribing and diagnosing events.

As an introductory example, consider the domain of cardiovascular disease (CVD) which is a leading contributor to chronic disease related deaths [1]. It is estimated that CVD affects around 80 million Americans (36.3% of total population), out of which 73.6 million (or 33.3% of total population) are estimated to have uncontrolled blood pressure (BP) [5]. The relationship between BP and risk of CVD events is continuous, consistent and independent of other risk factors – the higher the BP, the greater the chance of having a heart attack, heart failure, stroke and kidney disease [6]. Given the high prevalence and the CVD risk posed by uncontrolled BP, being able to identify patients who have had consecutively high BPs over time, or patients who have not received any antihypertensive medication for a defined period of time (say 90 days) after being diagnosed with hypertension presents an opportunity to intervene in patients whose clinical outcomes can be improved. Formulating these types of queries requires the explicit consideration of temporal relations; in this case between BP measurements, and diagnosis events and antihypertensive prescribing respectively.

1.2 The Opportunity

The uptake of information technology in the healthcare industry together with the use of electronic medical records (EMRs) has had a positive influence in improving healthcare
efficiency and safety. Nearly all the NZ general practices use some type of a specifically designed computer-based patient management system (PMS) to assist with recording of patient and clinical consultation details and to help with the daily running of their business [7]. Accordingly, a recent Commonwealth Fund/Harris Interactive survey [8] has placed NZ (out of 11 countries that were surveyed) in the top tier in terms of GPs using EMR systems, with the NZ systems rated as having the highest rate of providing most of the 14 desirable computer functions. My experience in analysing data from four NZ general practices shows that there are approximately 85 prescriptions written, about four patients diagnosed with some condition and about 20 BP measurements taken each day. General practice EMR data has been shown to have high completeness for prescribed items (99.7%) and laboratory tests (100%) [9]. Based on the premise that analysing prescribing data has the potential to reveal substantial gaps in therapy, the vast amount of high-quality, routinely collected EMR data provides a great opportunity to use this data for quality audit purposes of chronic disease management in accordance with evidence-based criteria.

My supervisor has an extensive track record of analysing EMR data [10-12] and at the time I started my PhD, my supervisor was beginning to put together a University based research team to collaborate with a general practice in West Auckland with a predominance of Pacific Island patients. I became a member of this team and the team conducted a series of sessions with the collaborating general practice. The primary focus of these meetings was to develop a set of quality indicators around chronic disease management, focussing mainly on hypertension and its common comorbidities. These quality indicators form the basis of indicators I have used as guidance for the requirements that the proposed framework should be able to formulate. Details of these meetings and the quality audit indicator development process are described in detail in Chapter 3. Note that quality indicators are commonly referred to as performance measures, clinical indicators and sometimes as audit criteria, and therefore these terms are used interchangeably in this thesis.

1.3 The Problem

Most, if not all, PMSs provide some type of ‘query-builder’ and/or reporting functionality, primarily for day-to-day statistical and/or financial reporting purposes. These built-in tools are usually designed to directly query for data in the underlying database tables with simple filters, write queries based on straightforward database table joins and also generate pre-defined reports. However, formulating queries related to chronic disease often requires the careful consideration of temporal relationships between different events which is difficult and not
always straightforward; also, query formulation often involves higher level concepts, such as 'hypertension', a concept that needs to be abstracted from raw data stored in a database, usually based on a selected standardised classification scheme. As such, there is a need for the development of novel query-formulation tools, which has been a key driver behind this research.

1.4 The Research Question

Given the background of this research, the primary research question this thesis aims to answer is whether a computational framework can provide quality audit of chronic disease management working from routinely collected EMR data? More specifically, can an architecture be developed that can (1) be configurable to address a wide range of temporal quality audit indicators, and (2) use appropriate technologies to provide easy maintenance of relevant clinical concept specification.

1.5 The Proposed Solution

The first attempt towards answering the research question and engineering the audit criteria was a somewhat straightforward database solution. A range of drawbacks was identified with this initial approach which led to a Semantic Web technology [13] based solution. The first approach was using an already optimised commercial database query execution engine while the second had the advantage of having sharable knowledge bases that are easy to maintain. This led to a hybrid solution which combines the independent querying and knowledge maintenance capabilities of the two previous approaches. The resulting final framework has been named ChronoMedIt, indicating a system for Chronological Medical audit, and much of this thesis is devoted to discussing the details related to the development and implementation of ChronoMedIt, as well as some of its applications.

The work presented in this thesis is a stepping stone towards improving the quality and uptake of quality audit criteria for process improvement in chronic disease management. My supervisor and I have focussed primarily on hypertension management although the proposed framework, techniques used and audit criteria are representative of the general nature of clinical concerns GPs would like to account for. During the research methodology development stage, we took the premise that our approach should be a clinician driven approach, which led to the discussions with the expert panel of currently practising clinicians to identify key requirements from a practice/practical perspective.
Our approach stands in contrast to trying to create a system that provides treatment recommendations (i.e., CDSSs). For example, the ATHENA [14] and PRODIGY [15] projects are arguably the most well known systems used for improved hypertension management; an underlying factor common to ATHENA, PRODIGY and most other decision support tools is that they all attempt to automate fully the relevant CPGs. Although almost all the systems have the capacity to let the GPs override the system recommendations and some have critiquing modes where the system interrupts only when critical alerts need to be made, we take the stance that auditing general practice with specific criteria provides a complementary capability for the management of patients with chronic disease. The value of our approach is further backed by the difficulties encountered in successfully integrating CDSSs into clinical practice, for instance, the problems reported during PRODIGY phase II implementation [15].

1.6 Thesis Outline

This thesis consists of eight chapters, and a brief description of each of the remaining chapters is given below.

Chapter 2: Related Work – This chapter is devoted to discussing a range of thesis-related literature. The main themes are discussing computer-interpretable CPG modelling techniques, guideline based CDSSs as relevant to chronic disease management, temporal analysis and visualisation tools and quality indicators in primary care.

Chapter 3: Development of Quality Audit Criteria for Chronic Disease Management – The main purpose of this chapter is to discuss some of the preliminary work that led up to my PhD research. It focusses on the audit criteria development process and demonstrates an initial implementation of a query-generation system. The chapter also discusses a field-validation study that was carried out to validate the importance of general practice prescribing data analysis for quality improvement efforts.

Publications related to this chapter are: [16-19]

Chapter 4: Preliminary Results – An Ontology Based Approach – Discussing a Semantic Web technology based implementation is the primary focus of this chapter. It discusses some of the limitations of current querying tools built into PMS systems and the importance of using a knowledge driven approach.

Publications related to this chapter are: [20-22]
Chapter 5: ChronoMedIt Framework Architecture – This chapter presents details related to the core implementation of the final framework architecture.

Publications related to this chapter are: [23-26]

Chapter 6: Framework Testing – Software testing is an important aspect of any software development effort. As such, this chapter is devoted to discussing some software testing techniques and how several testing techniques have been applied to verify ChronoMedIt’s implementation.

Publications related to this chapter are: [23]

Chapter 7: Applications using ChronoMedIt – ChronoMedIt has been used in several studies to identify chronic patients on suboptimal therapy. This chapter discusses several such applications along with some (preliminary) details of an ongoing feasibility study that is based on ChronoMedIt’s patient-identifying capabilities.

Publications related to this chapter are: [21, 23-25, 27-32]

Chapter 8: Discussion – This final chapter presents a discussion on the main contributions of this PhD research, ChronoMedIt’s limitations and the current status and possible future directions.
Chapter 2

Related Work

The previous chapter provided a high level view of the importance of the work presented in this thesis and discussed the need for new research. The purpose of this chapter is to establish a broad understanding of other related work. The chapter is divided into four main themes - (1) Computer Interpretable CPG Modelling Techniques, (2) Guideline based Chronic Disease Management CDSSs, (3) Temporal Analysis and Visualisation and (4) Quality Indicators in Primary Care.

The first section briefly discusses the importance of CPGs, the need to have them in a more computer interpretable format to provide decision support, and some of the important computer interpretable guideline formalisms that have been proposed to-date. The second section covers some of the important CDSSs for chronic disease that have been identified as being directly relevant to this work. The Temporal Analysis and Visualisation section describes the different approaches that explicitly consider the temporal dimension of chronic disease management. The last section is devoted to discussing the importance of using quality indicators in primary care for quality assurance purposes.

2.1 Computer Interpretable Clinical Practice Guideline Modelling Techniques

A multitude of CPGs have been developed by various national and international organisations mainly since the early 90s; however, active use of these CPGs within the day-to-day practice
environment is yet to materialise fully. Despite the lack of a universally agreed standard for quality of patient care, randomised controlled trials (RCTs) have shown that CPGs are an effective tool for improving the quality of care and for changing clinical practice [33]. The medical community in general accepts that following well developed and evidence-based CPGs improves the quality of care received by patients, along with the many other benefits CPGs provide [34, 35]. A widely used definition in the published literature for CPGs is that of the American Institute of Medicine, which defines a CPG as “systematically developed statements to assist practitioner and patient decisions about appropriate healthcare for specific clinical circumstances” [36] (several other definitions can be found in [4]). Systematically developing these CPGs is a complex task requiring much scientific rigour. The guideline development process usually includes steps such as composing the guideline development group to be representative of the key groups and disciplines affected, analysing the problem domain, formulating key questions, searching scientific literature, critically analysing the selected literature, evaluating and grading the evidence (using Delphi techniques [37] and/or a healthcare evidence grading tool such as the Grading of Recommendations Assessment, Development and Evaluation [GRADE] system [38]), developing evidence tables (to present the evidence behind recommendations), writing the draft guideline, external review by independent external referees and finally, guideline endorsement and publication [4]. The guideline development group is usually responsible for updating the guidelines as new scientific evidence/research emerges. The guideline development group also receives feedback from key groups involved in implementing the guidelines, including practicing clinicians. This guideline updating, maintaining and feedback loop continues in an iterative fashion with each iteration providing important (often incremental) refinements to the previously published version.

CPGs are developed with the intention of achieving a number of key goals: (a) minimising the variations observed in general practice and promoting consistency [7-9], (b) assisting and improving a clinician’s decision making process by reviewing the main options and providing recommendations [39], (c) optimising resource utilisation and outlay for hospitalisation and surgery [40, 41], (d) reducing healthcare costs [42] and (e) providing education, especially for less senior clinicians [36, 43]. Although CPGs are not the sole solution for ultimately improving clinical outcomes of patients [36], many countries including NZ, Australia, North America, Canada, United Kingdom (UK), Netherlands, Finland, Sweden, France, Germany, Italy and Spain [40] have shown increasing activity with CPG development, recognising their potential towards achieving optimal outcome for multiple stakeholders. There have been some
international level initiatives as well. For example, the Guidelines International Network (G-I-N) [44] is a recent international initiative that is involved in the systematic development of clinical practice guidelines and their application into practice, which it achieves by supporting international collaboration. As of September 2009, G-I-N has 93 organisations/partners (including a number of international organisations such as the WHO) representing 38 countries.

Research has established that passive CPG dissemination techniques such as publishing in journals and posting them to the clinicians are ineffective, have minimal impact and do not significantly change clinician behaviour [45-47]. Research has also shown that for CPGs to influence clinician behaviour positively, they need to be well implemented and deliver advice tailored to the individual patient’s needs at the point of care [34]. However, most CPGs are usually written as large documents in a textual format, and in order for a computer based system to be able to provide patient-specific recommendations based on best practice evidence-based CPGs, the CPGs need to be available in an electronic, computer interpretable format together with an electronic version of the patient profile. This requirement has given rise to the development of Computer Interpretable Guidelines (CIGs), which are essentially representations of the knowledge needed for a computer system to advise clinicians in a way that adheres to the guideline of interest [48]. It is highly desirable for CIGs to be sharable among institutions and a range of requirements for them to be sharable were identified by Boxwala et al. [49]. Advantages of a computer representable format of a CPG are that the guideline can be reviewed by specialist clinicians to ensure it contains the ‘best clinical practice’ knowledge, used to distribute up-to-date reference material rapidly and can be enacted by a computer to provide various recommendations [50]. Recognising this need for an electronic representation of CPGs for clinical decision support purposes, several groups around the world have developed guideline representation schemes with each having its own strengths and weaknesses. The remainder of this subsection is devoted to briefly discussing some of the important guideline representation mechanisms and their applications. The reader is referred to OpenClinical.org [51] which is a comprehensive source of information on all aspects of the different computer-interpretable guidelines developed to-date.

### 2.1.1 The Arden Syntax

The Arden Syntax is one of the best known early efforts designed for writing and sharing task-specific medical knowledge among healthcare providers [52]. It has a standard, Pascal-like syntax that can be used to define the logic for medical rules/decisions (e.g., rules for generating alerts for abnormal laboratory test results and drug interactions) without regard to the actual
implementation of that logic in existing systems. These rules are created as modular, independent rules, referred to as Medical Logic Modules (MLMs). An MLM is basically a text file that is arranged into discrete slots with different categories that enable MLMs to be more readable by humans. The ‘knowledge category’ (one of the MLM categories) contains the actual medical knowledge of the MLM and contains details on the way in which the MLM is used, rule preconditions, complex calculations and so on. The required logic is usually written in the form of ‘IF-THEN’ type statements [53]. The Arden Syntax has been used in several medical applications, such as the HELP system for providing generalised decision support, including alerts and reminders [54, 55], the CARE system to deliver reminders and hints to clinicians regarding patient treatment recommendations in an outpatient setting [53], as well as several others [56-58]. However, the Arden Syntax neither provides a strong underlying clinical performance model to guide application developers, nor offers declarative properties needed for effective and safe dissemination of expertise [50]. Maintenance of different MLMs has been proven to be a challenge [56], and since MLMs do not provide full support for conceptualising a multistep guideline, representing complex clinical guidelines with temporal and other relationship between tasks that often involve modelling alternative pathways (i.e., control flow) has been difficult [59].

Since the Arden Syntax, a number of new guideline modelling approaches have been developed that attempt to formalise CPGs. In two recent CIG analyses carried out by de Clercq et al. [60, 61], the process of developing guideline-based decision support systems was associated with four distinct areas: guideline modelling and representation, guideline acquisition, guideline verification and testing, and guideline execution. The researchers focussed on five CIG approaches as being the most important and refereed approaches. These approaches were the Arden Syntax (discussed previously), GuideLine Interchange Format (GLIF), PROforma, Asbru and EON. The latter four are discussed here briefly, but the reader is referred to [60, 61] for a comprehensive description on these five approaches.

2.1.2 GLIF

GLIF has been developed using a consensus-based multi-institutional development process informed by formal cognitive methodologies and contains many important features needed to represent CIGs [62]. GLIF models CPGs in terms of a flowchart that consists of structured scheduling steps, representative of the different clinical actions and decisions [60] and uses the Health Level Seven (commonly referred to as ‘HL7’) [63] Reference Information Model (RIM) for representing patient information [64]. The goal of GLIF is to enhance CIG shareability and
reuse across a variety of clinical settings, system architectures and institutions (a goal similar to that of the Arden Syntax) by modelling guidelines so that the CIGs are understandable by human experts as well as by the automatic parsers used in different CDSSs [60, 62]. The latest version of GLIF – GLIF version 3 (or GLIF3) [64] consists of an extensible object-oriented model and uses the resource description framework (RDF) [65] as the exchange syntax for guidelines. The model contains a number of classes that describe typical guideline characteristics, the attributes of those classes and the relationships among the classes that have been specified using Unified Modelling Language (UML) class diagrams. To model and represent a CPG, GLIF uses five steps: (1) decision steps to represent decision points in the guideline and direct flow control from one guideline step to other, (2) action steps for modelling recommended actions or tasks, (3 and 4) branch and synchronisation steps for modelling multiple concurrent paths through the guideline and (5) patient state steps to describe the clinical state that characterises a patient and to serve as an entry point into the guideline (depending on the current patient state). The GLIF encoded guidelines can be executed using the GLIF3 Guideline Execution Engine (GLEE) which also serves as an interface to the GLIF3 guideline representation model. GLEE provides defined interfaces to EMRs and other clinical applications to facilitate potential integration with local clinical information systems [66]. In order to specify decision and eligibility criteria, as well as abstract or derive summary values from GLIF encoded guidelines, the GELLO language has been specified which is a purpose-specific, object-oriented query and expression language [67]. A framework called the Knowledge-Data Ontological Mapper (KDOM) has also been developed to map guidelines encoded by GLIF3 to institutional specific EMR data items in an attempt to facilitate the process of CIG sharing [68].

### 2.1.3 PROforma

The PROforma CIG formalism has been developed with the intention of publishing clinical expertise in the form of expert systems which can assist patient care through active decision support and workflow management [50]. PROforma is based on a formal specification language and is formalised in terms of classical predicate calculus augmented by first order logic. When modelling a CPG in PROforma, the guideline is modelled as a set of tasks with data items. Instead of providing explicit support for various clinical domains (diagnosis, prognosis, treatment and so on), PROforma aims to provide a minimal set of abstract terms and properties which in turn can be adapted to suit various domains [69]. The task model in PROforma consists of four classes: actions to represent procedures to be executed (administering a drug for example), enquiries to represent points where information is to be requested (e.g., the age of a patient), decisions to model decision points, and plans which are essentially a collection of tasks
that are grouped together for logical reasons [70]. PROforma plans can include subplans (i.e., subtasks); thus the PROforma representation of a CPG leads to a hierarchical representation of tasks. All PROforma tasks items have several common properties alongside distinctive task-specific properties. The decision task has a 'candidates' property that is used to represent the different options that need to be considered when a decision is being made. Candidates have their own properties and can be associated with for (+), against (-) or irrelevant arguments to indicate the appropriateness of a particular candidate when making a decision. All PROforma tasks have precondition and task scheduling constraint properties to control the flow of guideline execution and also have four task-states which a task can be in: dormant, in_progress, discarded or completed [50, 70]. PROforma has two development environments, Tallis and Arezzo where Arezzo is a commercial version of Tallis. These development environments can be used to author a guideline using a two-step approach. In the first step, a high level diagram describing the temporal and logical sequence of the tasks is created using the graphical editor (composer). In the next step, this graphical structure is populated with detailed procedural and medical knowledge by entering the required information into the task properties attached to each task. The Arezzo or Tallis environments can also be used for guideline execution. These environments also provide a web interface which can be used to enact a guideline over the internet using a remote web server via Hypertext Transfer Protocol (HTTP) requests. These are processed by the hosting web server, calling a parser and the execution engine to enact the requested guideline. The web server may also be configured to obtain any required patient data from an external patient database server using connectivity protocols such as Java Database Connectivity (JDBC) [70]. PROforma is one of the few guideline representation formalisms that has made its way to the commercial arena and has been used on a wide scale in different clinical domains as the underlying CIG representation language; for example, the CAPSULE system was used to help GPs with prescribing decisions for common conditions by providing candidate medication with arguments for and against each recommendation, based on contraindications, drug interactions, side effects, costs and so on [69, 71]. Overall, PROforma has been used in seven studies (by 2006) that yielded quantitative data (and all have shown significant positive results on a variety of outcome measures) [71].

2.1.4 Asbru

Asbru is a time-oriented, intention-based skeletal plan-specification language used to represent and critique clinical guidelines and protocols [72, 73]. A skeletal plan is a plan schemata at various levels of detail that capture the essence of a procedure (i.e., similar to a template), but leaves room for execution-time flexibility in the achievement of a particular goal. Asbru views
CPGs as reusable, time-oriented skeletal plans that need to be refined over significant time periods; therefore, when a plan needs to be executed on a given patient, the corresponding skeletal plans are instantiated. An Asbru plan consists of five components: preferences which indicate bias or constrain the selection of a plan to achieve a given goal; intentions which represent high-level goals (e.g., patient BP should never exceed 140/90 mmHg); conditions which specify constraints, such as plan preconditions; effects which describe the functional relationship between plan arguments and measurable parameters, such as ‘dose of insulin is inversely related to blood glucose level’; and a plan body which is a set of subplans or actions that need to be executed in parallel, in sequence, in any order, or in some frequency (e.g., cyclical) when the plan is considered appropriate [72]. In order to manage these often complex skeletal plans, Asbru is also able to specify temporal patterns (e.g., the concept of HIGH for a given parameter over a certain time period) and time annotations (to represent uncertainty) as well as user interfaces to visualise the developed plans. Several tools have been developed to assist the process of creating an Asbru encoded representation of a CPG. Votruba et al. [74] suggest that the bottleneck in the clinical guideline and protocol development cycle is the acquisition, validation/verification and maintenance of guidelines, and they propose the Document Exploration and Linking Tool/Addons (DELT/A) to support the translation of HyperText Markup Language (HTML) documents into a formal representation in Extensible Markup Language (XML). DELT/A allows a protocol designer to establish links between the original guideline and its computerised version and has been successfully used to edit Asbru protocols. AsbruView is a visualisation tool and also a user interface that is designed to be suitable for physicians and uses graphical metaphors, such as a running track with a finishing flag, to represent Asbru plans [75]. Another Asbru related tool is CareVis, which has been developed to integrate and combine classical patient data visualisation with the visualisation of treatment information in terms of logic and temporal aspects. CareVis provides multiple views of Asbru plans, with a Logical View for plan visualisation and a Temporal View that focusses on the time-oriented aspects of Asbru plans, as well as the representation of parameters and variables over time [76]. Asbru CIGs have been used in several practical applications as well; for example, the Oncocure project looks at developing a prescriptive guideline-based CDSS to provide active support at important decision steps by executing Asbru-encoded protocols of several pharmacological therapies for breast cancer. This project has investigated ways of integrating the Asbru interpreter with the existing clinical database and the graphical user interface (GUI) of the EMR in an effort to reduce costs of care while improving adherence to guideline and quality of documentation [77].
2.1.5 EON

More of a framework than a typical CIG formalism, EON [78] was developed with the intention of promoting a modular architecture that encapsulates potentially reusable and extensible functionality into discrete software components. This type of architecture makes it easier to modify and maintain the associated software as the changes are contained to the individual modules, thus reducing development and maintenance costs. These modules are based on problem-solving methods that are those implemented in software to perform some conceptually abstract goal, such as generating a plan given a set of constraints, or generating an abstraction given some raw data. They provide a standardised computational algorithm that can be reused in a variety of contexts, but usually differ from an easily understood and well-defined subroutine type functionality where the input-output relationship can be (easily) specified, for example, via a direct mathematical formula. EON consists of three key knowledge representation modules [79]: the patient data information model where the classes and various attributes of patient data are defined as relevant to the guideline; the medical-specialty model which contains medical domain-specific knowledge that models the structure of domain concepts, such as drugs and treatment; and the guideline ontology which conceptualises a particular guideline. EON aims at modelling multi-encounter patient management, such as chronic disease management, where a patient’s state is associated with a number of decisions and actions. To cater for this requirement, the core guideline ontology consists of four important classes: scenarios to represent patient states, decisions to model choices among alternatives, actions to represent acts that lead to changes in the state of the world (such as collecting patient data) and goals to formally specify a guideline on a higher level of abstraction [60, 61]. Although EON is no longer an active project, it is currently being used within the ATHENA system (see Section 2.2.1), a CDSS that encourages BP control and recommends a guideline-concordant choice of drug therapy [14]. The EON framework has strong temporal abstraction support via its modular components and more details are discussed in Section 2.3.2.

2.1.6 CIG Review and Comparison Studies

The medical informatics community has shown a lot of interest in the development of CIG representation formalisms, especially during the 90s. As a result, a number of CIG representation formalisms have been developed other than those discussed thus far. Among the other important formalisms are GASTON [80], GLARE [81], GUIDE [82], HELEN [83] and SAGE [84]. Each of these representations has its own significance within the CIG community, but as de Clercq et al. [60, 61] indicate, these other formalisms have not had the same impact in terms of publications and clinical applications.
Various studies have compared the similarities and differences among a wide range of CIG representations (often along a selected aspect of modelling). Tu et al. [85] analysed a range of graphical representations that are used by CIG modelling methods to express process information that is embodied in clinical guidelines and protocol. This analysis identified a typology of four modelling formalisms used by guideline models: (1) flowcharts for algorithmic problem-solving processes, (2) disease-state maps that relate various decisions made during the course of patient care, (3) sequencing of activities in care plans that aim to meet specific goals and (4) workflow specifications that model care processes in organisations. Peleg et al. [59] suggest that although each CIG model has adopted a different approach based on developer interests and expertise, many of the CIG representations share a “hierarchical decomposition of guidelines into networks of component tasks that unfold over time” and typically provide modelling primitives designed for modelling complex, multistep clinical guidelines that can be used to describe temporal and other relationships between various tasks. These types of modelling formalisms have been referred to as ‘Task Network Models’. Six different CIG models – Asbru, EON, GLIF, GUIDE, PRODIGY and PROforma – were compared in [59] using a case study based approach where CIG models were created for two guidelines (a chronic cough related guideline and a hypertension guideline), using the six representations. The different models were then compared according to eight components that were needed to encode a guideline and for an encoded guideline to be linked with patient data. The study found a consensus on many components; the differences were primarily in the underlying decision models, goal representation, use of scenarios and structured medical actions. de Clercq et al. [60] compared five approaches – the Arden Syntax, GLIF, PROforma, Asbru and EON – in terms of their representation, acquisition, verification and execution aspects and concluded that design specifications made in one area (such as guideline representation) have implications in other areas (such as guideline execution) and that several issues related to guideline implementation and guideline-based decision support still have to be addressed more extensively, although significant progress has been made in guideline representation techniques. Asbru, EON, GLIF and PROforma were evaluated with a focus on their control-flow features by Mulyar et al. [86] where it was found that all four directly support basic control-flow patterns, cancellation patterns and some advanced branching and synchronisation patterns. Wang et al. [87] studied 11 CIG representation models along representation primitives, process models (i.e., scheduling constrains/temporal order on representation primitives and nesting of guidelines during guideline application) and modelling of patient data. Decisions and actions were identified as the key representation primitives that should be provided by a
guideline representation model while sequences, concurrency, alternatives and loops were identified as the important scheduling constraints. The researchers found little consensus on a common patient data model however, and point out how a critical requirement to have CIGs applied in clinical practice is the integration of the CIG data model with EMRs, and ideally with physician order entry systems. CIG related literature from 2000-2007 was reviewed by Isern and Moreno [88] where eight systems with guideline execution engines were selected and compared along language used and the architecture of the system. The authors concluded that although “these systems could be beneficial for clinicians and patients, it is an ongoing research area, and they are not yet fully implemented and integrated into existing careflow management systems and hence used in daily practice in healthcare institutions.” This appears to be a fair reflection of the current status of CIGs with respect to their use in clinical practice, even three years later.

All the CIG modelling formalisms discussed so far offer generic guideline modelling constructs for representing careflow processes as a network of tasks. These constructs are used as ‘building blocks’ to formulate the computer-interpretable representations of guidelines to convey the various medical decisions and perform relevant actions. Peleg and Tu [89] propose a different approach where the guideline itself can be abstracted into patterns modelling clinical knowledge that conform to sets of predefined templates. These patterns are independent of a particular guideline modelling language and offer templates that can be used to structure clinical knowledge in a top-down manner, thus breaking the clinical knowledge into manageable interacting components. Clinical templates have been developed and validated with respect to screening guidelines and immunisation guidelines while taking temporal constraints into account; however, recurring patterns (such as those that commonly occur in chronic disease management guidelines) are yet to be developed. Furthermore, encoding of indication and contra-indication constraints is not yet supported within the templates; if these need to be modelled, the templates need to be complemented with some other ‘expression language’ that provides such constraint support.

As discussed already, due to the relatively independent efforts by research groups developing CIG formalisms, a number of CIG representations have been proposed over the past years. Developing a formal representation of a guideline is desirable, yet the development process is often difficult and expensive. Furthermore, the formal model is usually difficult to compare to the original guideline. Seyfang et al. [90, 91] indicate that this is due to the large gap that exists between natural language and the currently available formal representations. In an attempt to bridge this gap, the researchers propose an intermediate representation called a Many-Headed
Bridge (MHB) between informal representations such as free-text and tables, and the more formal representations such as Asbru, GLIF or PROforma. Although the degree of formalisation varies between each of the CIG approaches, the majority represent a guideline in a format that is precise enough to be executed (semi-) automatically, and hence an MHB type representation is feasible. An MHB is essentially a small and versatile ontology of guideline components that groups together the statements in the guideline into ‘chunks’ with predefined dimensions such as control flow, data flow, temporal aspects and evidence. MHB representations developed for Asbru have been shown to be appropriate to model important guideline statements, easier to create an MHB model from the original guideline text than an Asbru model, and easier to create an Asbru model from the MHB than from the original text [91].

2.1.7 LASSIE and DeGeL

An aspect that warrants consideration (irrespective of the CIG representation formalism) is presenting up-to-date and state-of-the-art medical knowledge to practitioners. New research findings are continuously published and the CPG development process often takes at least two years; as such, CPGs can be out of date as soon as they are produced. This issue has been traditionally addressed by specifying a shelf-life by a date (e.g., “this guideline will be reviewed in two years”) or by a statement that the review date will be determined by the availability of new evidence (e.g., “this guideline will be considered for review as new evidence becomes available”). As a solution, Kaiser and Miksch [92] introduce the notion of a ‘living guideline’ which is a guideline that remains under review on an ongoing basis, with updates published at set intervals (annually, for example). The researchers propose a semi-automatic methodology called LASSIE that can be used to formalise guidelines in any guideline representation language by applying a two-step process: first applying general steps (that are language-independent) and then language-specific steps. The LASSIE methodology uses information extraction techniques to semi-automatically obtain as many fragments of formal representation from a guideline in free-text as possible and provides methods for comparing and matching different guideline versions. LASSIE also provides linking between text parts of the CPG and their corresponding formal models; as such, the unchanged parts of the previously formalised CIG can be easily identified. As a result, only the updated text parts of a CPG need to be formalised when the medical knowledge needs to be updated, thus reducing the effort required in the CPG formalisation process.

Other than maintaining and updating guideline knowledge bases, it has been suggested that a primary hurdle preventing fast conversion of free-text guidelines into computer interpretable
formats is because expert physicians cannot program using the syntax of the formal guideline specification languages while, conversely, programmers and knowledge engineers do not understand the clinical semantics of the guidelines; this is referred to as the “guideline conversion problem” [93]. The text-based representations have their own advantages, primarily for search and retrieval of relevant guidelines, although a formal representation is necessary for them to be used in decision support. Starting from a free-text CPG, Shahar et al. [93] propose three main steps to be followed during a guideline conversion process: (1) expert physicians (i.e., the domain experts) classify the guidelines along multiple semantic axes, and semantically ‘markup’ (i.e., label portions of the text by the semantic labels) the existing text-based guidelines, thus creating a semi-structured format (which is still text-based); (2) the expert physicians further structure the guideline, possibly with the assistance of a knowledge engineer, into a semi-formal structure, which includes ontology-specific control-flow knowledge; and (3) knowledge engineers convert the marked-up text or the semi-formal structure into a formal, fully structured and computer-interpretable representation of the target ontology using an ontology-dedicated tool. The knowledge in the text-based CPGs is often implicit in nature, clear only to the expert physician authoring the guideline, and must become explicit during the conversion process. As such, this type of conversion process has the advantage of capitalising on the relative strengths of both expert physicians and knowledge engineers. The Digital electronic Guideline Library (DeGeL) [93, 94] has been proposed to facilitate this gradual conversion process. DeGeL has been implemented as a web-based, modular, distributed architecture with a set of associated software tools for incremental conversion of guidelines into multiple guideline-specification representations. These tools are used for performing tasks such as guideline specification, semantic markup, search, retrieval, visualisation, eligibility determination, runtime application and retrospective quality assessment [94]. The main tools that are associated with DeGeL are:

- URUZ [93] – a web-based guideline markup tool that enables medical experts to create new guideline documents,
- Vaidurya [95] – a generic guideline search engine providing search (full-text, concept-based and context-sensitive) and retrieval features,
- VisiGuide [93] – a browsing and visualisation tool that enables users to browse the guidelines returned by the Vaidurya search engine and visualise their structure,
- IndexiGuide [93] – a tool that supports the semantic classification of a new guideline being uploaded to the DeGeL library to facilitate later retrieval,
- the Spock system [96] – a runtime environment that supports the application of guidelines at various levels of formatting,
- QualiGuide [93] – an assessment tool to evaluate clinician adherence to guidelines, and
- GESHER [97] – a graphical guideline acquisition tool that manages the incremental guideline specification process.

To overcome some of the limitations of the web-based architecture of DeGeL, a newer version is currently being developed, called DeGeL.NET [98] which is a distributed, web-service based, open architecture implementation based on the Service Oriented Architecture (SOA) design specification.

2.2 Guideline based Chronic Disease Management CDSSs

CDSSs are information systems designed to directly assist and improve the clinical decision making process by providing clinicians with patient-specific assessments or recommendations that are based on matching the appropriate characteristics of an individual patient to a computerised knowledge base [99, 100]. Among other positive influences, CDSSs have been shown to improve clinicians’ prescribing practices, reduce serious medication errors, enhance the delivery of preventive care services, improve adherence to recommended care standards and increase healthcare quality and efficiency [101]. Ideally, all CDSSs could be based directly on an underlying CIG since the promise of CDSS–facilitated evidence-based medicine is strong, however, substantial work still remains to be done to fully realise their potential benefits [100].

CDSSs are commonly used across almost all the disciplines in medicine, but due to the scope of this thesis, only a few important CDSSs with a focus on chronic disease management are considered. This section will briefly describe these CDSSs that have been identified as directly relevant, based on academic literature searches, domain expertise and the popularity of various systems, and does not attempt to provide a comprehensive review of all chronic disease management CDSSs.

2.2.1 ATHENA

The ATHENA DSS (Assessment and Treatment of Hypertension: Evidence-Based Automation Decision Support System) is a point-of-care decision support system that presents recommendations to clinicians for the treatment of hypertension based on widely accepted national guidelines for hypertension, particularly the Sixth Report of the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure (JNC6) [102]. The ATHENA DSS serves as a reminder system to encourage clinicians to focus on
achieving adequate BP control and also provides the more sophisticated and complex reasoning necessary to operationalise JNC6 and thus arrive at therapeutic recommendations [14]. The DSS can make recommendations to add, substitute, delete or change the dose of drugs on the basis of how well BP is controlled. It can also accurately account for comorbid conditions and recommend compellingly indicated therapy (per JNC6), for example, recommending beta blockers after a myocardial infarction. Other categories of therapy to be considered include possible indications (e.g., thiazides for patients with osteoporosis), possible clinical concerns (e.g., beta blockers for patients with depression), contraindications (e.g., beta blockers in the presence of asthma) and allergy/previous adverse drug event information. All recommendations are provided with a rationale for the addition, substitution, deletion or dosage change, as well as pointers to the guideline step that is most relevant to the patient's current condition. The ATHENA DSS has been implemented as a platform-independent system and uses patient data from the Department of Veterans Affairs’ medical record system (called VistA – Veterans Health Information Systems and Technology Architecture) to provide recommendations [103]. ATHENA is currently being used at multiple sites at the Department of Veterans Affairs in the US [104] and provides recommendations to clinicians via a popup window (Figure 2.1) that is displayed on top of the existing EMR system’s GUI [14, 103, 105].

![Figure 2.1: A sample ATHENA advisory recommendation appearing as a popup window on top of the existing EMR system [106]](image-url)
ATHENA also has some graphical displays to summarise and clarify complex interrelationships. For example, ATHENA provides a BP-Prescription Graphs tab (Figure 2.2) to display all the BP readings and antihypertensive agents for the selected patient on the same timeline, with a scroll bar to view older data [103].

![Figure 2.2: A sample BP-Prescription graph produced by the ATHENA DSS [106]](image)

The ATHENA DSS has two main components: a knowledge base that models hypertension knowledge independently of its use, and a guideline interpreter that combines patient information from the EMR with the knowledge in the knowledge base to provide patient-specific treatment recommendations, explanations and evidence-based education that are consistent with the guideline recommendations [103]. ATHENA has been constructed using the EON architecture (see Section 2.3.2) which consists of a knowledge base that models the hypertension knowledge, a guideline interpreter and a temporal database mediator to support (temporal) querying. The knowledge base consists of inclusion and exclusion criteria, rules for hypertension control with separate rules for diabetics and non-diabetics and drug therapy logic and recommendations (including those for comorbid conditions and a history of adverse effects). The knowledge base is maintained separately from the logic involved, allowing clinicians to maintain and modify the ATHENA knowledge base as medical policies are changed and/or locally elaborated. The underlying architecture of ATHENA is EON which is a general architecture for guideline-based decision support. As such, ATHENA uses EON’s features to determine whether a guideline is applicable to a patient, the portions of the guideline that are applicable, satisfaction of guideline goals (e.g., has goal BP been reached?) and also apply criteria for selecting one course of action over another when generating advisories
about therapy [51]. ATHENA has been used in clinical settings to determine barriers for clinician adherence to hypertension guidelines [107, 108]. As a result of sufficiently addressing various issues identified through testing [109] and addressing socio-technical aspects related to implementation, ATHENA implementation had gone well and the system had run for more than 15 months (by 2004) in clinic sites in nine different cities that were part of three administratively separate medical centres with ATHENA being used to display detailed, individualised advisories for more than 10,000 patients [103].

2.2.2 PRODIGY

PRODIGY (Prescribing RatiOnally with Decision-support In General-practice studY) was a guideline-based decision-support system funded by the National Health Service (NHS) in the UK that was developed to assist GPs with choosing rational therapeutic actions for their patients [15]. The PRODIGY project was launched in 1995 as a three-phase iterative study, with the evaluation findings of each phase being used to refine the next stage of the project [110]. PRODIGY Phase I and II systems were implemented by commercial primary care information system vendors in the UK as modules that extended their existing EMR systems, using a similar look-and-feel to those proprietary systems, and accessing the coded information in the EMR to help direct choices in the guidelines. The PRODIGY system includes a guideline model, which in PRODIGY II was used to implement guidelines for the management of acute diseases [15]. PRODIGY Phase I was evaluated in 1996 and Phase II in 1998, both with encouraging results [110]. Phase II showed significant improvements over Phase I in terms of usage and patient information leaflets (these are leaflets containing general information about the associated disease in non-clinical terms) [111]. PRODIGY II was found to be technically competent at providing advice for acute disease management, shared decision making and prescribing. However, clinicians judged that the representation of guidelines for chronic diseases was inadequate and guideline authors found that creating guidelines for chronic disease management in PRODIGY to be a difficult task with unsatisfactory results. PRODIGY Phase III was initiated to address these issues and to develop a system that clinicians and guideline authors found more intuitive [15].

PRODIGY Phase III contains a Virtual Medical Record (Figure 2.3) that consists of eight classes of objects with a corresponding set of ‘PRODIGY III terms.’ Three of these objects identify the Patient, Care Provider and an Encounter while the other five are related to identifying the type of record [112].
Various queries and records to be written to the EMR need to conform to this Virtual Medical Record model and should be coded using PRODIGY III terms. As a result, all supplier systems need to provide terminology mapping services (using mapping components) to convert the vendor-specific native terminology into PRODIGY III terms, and vice-versa. The PRODIGY execution engine specifies a number of interfaces and services that system vendors need to implement on their systems to facilitate tasks such as querying the EMR and performing actions (‘prescribe beta-blocker’ for example). The execution engine itself also provides a number of interfaces and services so that vendors can use them to implement the user interaction modules within their proprietary system using the system-specific look and feel while maintaining the same user-interaction/functionality across vendors [112].

PRODIGY III models a guideline as a high level view of a network consisting of scenarios, action steps and subguidelines (Figure 2.4).
A scenario is an easily recognisable patient state for a particular diagnosis such as 'hypertensive on triple therapy' or 'angina on non-pharmacological treatment' (denoted by ovals in Figure 2.4). Scenarios provide an intuitive set of access points into a guideline so that a clinician can easily position a patient within the guideline depending on the state of the patient. This type of ‘manual positioning’ feature within a guideline was determined to be suitable based on the premise that EMRs are often incomplete, and therefore the system cannot be expected to always position the patient accurately in the guideline. An action step is a potential choice of management for a given scenario and outcome assessment that is used to make simple management decisions. Decisions are based on rule-in and rule-out conditions that are associated with each available alternative to determine the preferred course of action. Subguidelines are used when a single action step is not sufficient to describe the intended process of care, and further decisions and rules are necessary to refine the chosen action step(s) (which in turn can lead to further scenarios and action steps to be assessed within the consultation) [15]. In essence, the management of a patient in PRODIGY III over time according to guideline specifications can be viewed as the traversal of a number of selected scenarios and associated actions and further decision points along a single path [113].

Compared to ATHENA (discussed previously), PRODIGY III takes the view that providing some flexibility in guideline use is important from a clinician’s perspective (as opposed to detecting the patient’s current state automatically and then providing decision support). The PRODIGY III model provides documented guidance on the available choices and lets the clinician select one of the probable scenarios to control the main decisional step and explore the list of therapeutic choices, eventually browsing any associated documentation [113]. Guidelines related to chronic illness, especially for asthma, hypertension and stable angina have been implemented in PRODIGY III. However, evaluation of the asthma and angina management implementations involving 60 British general practices found no effect in adults in primary care [114]. This lack of effect was possibly due to low levels of software use despite the system being optimised as far as was technically possible. A subsequent qualitative study with five general practices identified three main areas of concern among clinicians; timing of the guideline trigger, ease of use of the system and helpfulness of the content [115]. It has been indicated that even if the technical problems of producing a system that fully supports the management of chronic disease were solved, the challenge would lie in addressing the issues encountered when integrating a system into clinical encounters where busy practitioners manage patients with complex, multiple conditions [114].
Although the PRODIGY project has been discontinued, the PRODIGY knowledge has evolved into the NHS Clinical Knowledge Summaries (CKS) where the knowledge is used as a reliable source of evidence-based information and practical ‘know how’ about common conditions managed in primary care. CKS are aimed at healthcare professionals working in primary and first-contact care and provide quick answers to real-life questions that arise during consultations, linking to detailed answers that outline the evidence on which they are based [116].

2.2.3 ASTI Project

The ASTI project [117] is a French national research project that aims to develop a decision-support system to support prescribing in primary care. The system has been developed to provide guideline-based support for prescription writers to prevent two types of medication errors: (1) incorrect use of a drug or an inappropriate combination of drugs (incorrect dose, drug interactions or contraindications for example); and (2) inappropriate choice of a therapeutic strategy (e.g., non prescription of the most appropriate drug). The system is integrated into existing EMRs and attempts to prevent errors of the first type through the use of the French national drug database (called ‘Banque Claude Bernard’) designed to check for possible issues; errors of the second type are avoided by providing guideline information to improve therapeutic strategy recommendations. This leads to two modes of operation, a critiquing mode and a guided mode. The critiquing mode operates as a background process that monitors prescribing and querying actions and proposes alternative medication if better ones are available and warns the user only when obvious strategic errors are detected. The guided mode on the other hand needs to be initiated by a clinician and is used to search for the best patient-specific treatment recommendation. This mode can provide interactive browsing of the knowledge base permitting contextual interpretation of the clinical situation to ensure that the solution reached is indeed the best patient-specific solution.

The first phase of the ASTI project has led to a prototype developed to assist GPs with prescribing for hypertension; however, extensions to other diseases such as type-2 diabetes are in progress. The ultimate goal of the project is to extend the approach to various other diseases to promote the use of therapeutic knowledge bases and their associated tools to be incorporated into existing EMR systems in primary care [118]. Preliminary evaluations of ASTI have been carried out (using a before-and-after methodology) to validate the hypothesis that GPs would use the critiquing mode for simple cases and the guided mode for more complex cases. The results indicated (as expected) that reminder-based interaction (i.e., the critiquing mode) is
appropriate for simple cases and that physicians are willing to use the on-demand system (i.e., the guided mode) as clinical situations become more complex, making the two modes of ASTI complementary [119].

2.2.4 HyperCritic

HyperCritic is a CDSS that solely relies on the information already in the EMR (as opposed to prompting for information as in a ‘consultation’ type model) and attempts to offer comments to GPs on their treatment/management of hypertension. The system relies on the GP to have correctly recorded the initial hypertension diagnosis into the EMR. For diagnosed patients with hypertension, HyperCritic can critique on four different aspects of care – preparation (for actions that need to be done prior to performing an action), selection (to report on inappropriate actions, such as prescribing beta-blockers in the presence of asthma), monitoring (to monitor subsequent activities such as observing a particular patient parameter after an action) and responding (to report findings related to an action, such as increased frequency/severity of asthma attacks after prescribing a beta-blocker) [120, 121]. HyperCritic has been evaluated primarily within two clinical settings – (1) 20 patients with hypertension were assessed by eight physicians as well as HyperCritic with the finding that out of 260 comments made by physicians, 118 were also made by HyperCritic; a range of limitations around failing to produce the remaining 142 comments were identified with insufficient data in the EMR, absence of sufficient medical consensus and recording clinical information as free text being among the important limitations [121, 122]; (2) 20 EMRs describing in total 243 visits of patients with hypertension were assessed by eight reviewers (two internists and six GPs) and by HyperCritic. A panel of 14 GPs subsequently analysed the relevance of those critiques using a five-point scale ranging from relevant critique to erroneous or harmful critique. 65% of HyperCritic’s comments were determined to be relevant or somewhat relevant, while this was 61-68% for the reviewers’ comments. Large differences were observed among panel members, further indicating the need for sufficient medical consensus [123].

2.2.5 AsthmaCritic

AsthmaCritic [124] is a CDSS that is used to review a clinician’s treatment of asthma and chronic obstructive pulmonary disease (COPD) in the light of the most recently published guidelines. The system is a non-inquisitive critiquing system that solely relies on EMR data and does not ask for any additional data entry. AsthmaCritic becomes activated by the EMR at the end of each patient encounter where it first searches the EMR for clues and triggers that indicate the possibility of asthma or COPD, based on International Classification of Primary Care (ICPC)
codes [125] for asthma or COPD diagnosis and Anatomical Therapeutic Chemical (ATC) codes for prescriptions used in the treatment of asthma or COPD. When a trigger is encountered, the record is selected for further analysis, and is subsequently reviewed for various aspects of treatment. AsthmaCritic can generate patient-specific feedback in the form of critiquing comments and distinguishes between two different kinds of feedback: critique and transformed clinical measurement. Critiques are patient-specific comments based on the patient’s current clinical situation, where a brief overview of all comments ranked by clinical urgency and potential impact are displayed to the clinician. Transformed clinical measurements based on current patient measurements and various calculations are presented to the clinician (that can then be printed and given to the patient). For example, peak-flow values may be presented in an overview that includes the patient’s values alongside the expected value for the patient based on gender, age and height, and also the original values expressed as a percentage of the expected value [126]. Both types of feedback review the physician’s diagnostic and therapeutic interventions, thereby enabling clinicians to reflect on their current decisions. For each comment, the system is able to provide additional information, such as more elaborate advice, further explanations and the relevant patient data as well as the underlying medical reasoning, including references to the source material on which the comments are based (e.g., references to the appropriate sections of the guidelines) [124, 126].

AsthmaCritic’s underlying medical knowledge base was derived using the asthma and COPD guidelines of the Dutch College of GPs, guidelines on interactions and side effects, pharmaceutical reference books and the existing consensus among a group of specialists in asthma and COPD. AsthmaCritic was designed with the intention of directly integrating the system into daily practice and explicitly considered aspects of practical implementation issues, such as treating the user as a professional and limitations of a busy practice [126]. As a result, AsthmaCritic has been well accepted and a retrospective study with 28 Dutch practices showed that AsthmaCritic was able to detect asthma and COPD patient records for further analysis using routinely recorded data only [124]. A larger RCT in the Netherlands of 32 practices with 40 GPs (16 practices with 20 GPs were the control group) assessed the effect of AsthmaCritic on the monitoring and treatment of asthma and COPD in daily practice where the intervention group was given the use of AsthmaCritic while the control group continued working in the usual manner (both groups had access to the asthma and COPD guidelines routinely distributed by the Dutch College of GPs). The study found that the AsthmaCritic critiquing system changed the manner in which the clinicians monitored their patients (e.g., an increased
number of pulmonary function assessments and peak-flow measurements were conducted in the study group) and also their data-recording habits [127].

2.3 Temporal Analysis and Visualisation

The notion of time plays an important role in the medical domain. Temporal analysis has been a topic of great interest to researchers for over three decades and has its origins in the fields of computer science and artificial intelligence. Temporal analysis covers a broad spectrum of concepts in the medical domain and four significant areas of research have been identified [128]: (i) fuzzy logic, time and medicine; (ii) temporal reasoning and data mining; (iii) health information systems, business processes and time; and (iv) temporal clinical databases. Two of the most influential contributions that laid the foundation for most temporal formalisms are from Allen [129] and McDermott [130] where formalisms for reasoning about time intervals and points have been proposed based on temporal logic. The work by both scientists is rather theoretical/conceptual in nature (e.g., Allen attempts to propose theory that is powerful enough to capture the range of the meanings and distinctions expressible in English related to the concepts of ‘actions’ and ‘events’ [129]). However, the thirteen relationships between time intervals that Allen proposes (Figure 2.5), commonly referred to as the ‘Allen Operators’, are of special interest to this thesis. Figure 2.5 shows these relationships – all except the ‘equal’ operator will result in a different relationship when $X$ and $Y$ are interchanged (‘$Y$ before $X$’ for example), yielding thirteen operators.

![Figure 2.5: The thirteen possible relationships between time intervals. $X$ and $Y$ denote time intervals (modified from [129]).](image-url)

<table>
<thead>
<tr>
<th>Relation</th>
<th>Pictoral example</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$ before $Y$</td>
<td>XXX YYYY</td>
</tr>
<tr>
<td>$X$ equal $Y$</td>
<td>XXX YYYY</td>
</tr>
<tr>
<td>$X$ meets $Y$</td>
<td>XXX YYYY</td>
</tr>
<tr>
<td>$X$ overlaps $Y$</td>
<td>XXX YYYY</td>
</tr>
<tr>
<td>$X$ during $Y$</td>
<td>YYYY YYYY</td>
</tr>
<tr>
<td>$X$ starts $Y$</td>
<td>XXX YYYY</td>
</tr>
<tr>
<td>$X$ finishes $Y$</td>
<td>XXX YYYY</td>
</tr>
</tbody>
</table>

Figure 2.5: The thirteen possible relationships between time intervals. $X$ and $Y$ denote time intervals (modified from [129]).
This type of work was the foundation for temporal reasoning systems such as the RÉSUMÉ system, and also temporal querying systems such as Chronus, both discussed in some detail later in this section. However, a detailed discussion on theoretical/conceptual temporal analyses based on temporal logic is beyond the scope of this thesis. Therefore, in this section, only temporal analysis and visualisation tools with some practical implementation are discussed. The reader is referred to the methodological review by Augusto [131], the non-exhaustive overview by Combi and Shahar [132] and the position paper that discusses promising directions of research in temporal representation and reasoning in medicine by Adlassnig et al. [128] for further treatment of the various other research efforts related to temporal analysis.

2.3.1 Knowledge Based Temporal Abstraction and the RÉSUMÉ System

From a broad perspective, temporal abstraction refers to creating high-level summaries from time-oriented data. Creating such summaries is necessary because clinical databases usually store raw, time-stamped data whereas treatment protocols often require information at a higher conceptual level. Shahar [133] views the temporal-abstraction task as “a type of a generic interpretation task: Given a set of time-stamped data, external events, and abstraction goals, produce abstractions of the data that interpret past and present states and trends and that are relevant for the given set of goals”. In the clinical domain, a final diagnosis is not always the main goal and a coherent intermediate-level interpretation of the relationships between data and events, and among data within a given context (such as a diagnosis) may be the interpretation needed. As such, the goal of the temporal abstraction task is to abstract the data into higher-level concepts and create from time-stamped input data, interval-based temporal abstractions and more complex patterns, often involving several intervals. The temporal-abstraction task uses an inference structure and related-required knowledge that are specific to abstracting higher-level concepts from time-stamped data, but independent of any particular domain [133]. The knowledge-based temporal-abstraction (KBTA) method is essentially a knowledge-level representation of the temporal-abstraction task with the knowledge required to solve that task and is a general problem-solving method for interpreting data in time-oriented, knowledge-intensive domains. KBTA is a modular approach with clear semantics for both temporal abstraction and for specifying the domain specific knowledge needed by the abstraction process [134]. The KBTA method has a formal model of input and output entities, their relations and the domain-specific properties that are associated with these entities - i.e., the KBTA ontology (refer to [133, 134] for details of the KBTA ontology).
RÉSUMÉ is a system that performs temporal abstraction of time-stamped data [135] and implements the KBTA method [134] to solve the temporal-abstraction task [133]. RÉSUMÉ is composed of a temporal-reasoning module, a static domain knowledge base (i.e., the domain’s temporal abstraction ontology) and a dynamic temporal fact base (i.e., the input and output data) [134, 135]. The original temporal reasoning module was implemented in CLIPS, a rule-based expert system shell [135]. RÉSUMÉ’s temporal-abstraction ontology consists of four knowledge structures to make domain-specific inferences on time-oriented data: a parameter-properties ontology to define the relevant domain parameters; an event ontology to define the events that can affect the interpretation of data; a context ontology to define the relevant interpretation contexts; and a set of dynamic induction relations of context intervals to relate events and parameters to the interpretation contexts that they induce [136]. The temporal fact base is used to loosely couple RÉSUMÉ to an external database (where primitive patient data and clinical events are stored and updated externally) as well as to store inferred abstractions and the input data [135]. A truth-maintenance system is used to handle any effects of updates made to the input parameter and event intervals which may cause the abstractions to be updated [134, 135].

The typical input to RÉSUMÉ would be a set of time-stamped parameters (such as laboratory measurements) and events (such as medical interventions) while the output consists of a set of abstractions. For example, given a set of raw haemoglobin levels, RÉSUMÉ can infer abstracted levels of haemoglobin within a specified context (see Figure 2.6). An important feature of RÉSUMÉ is that an abstraction of a parameter is also itself a parameter (I1 to I6 in Figure 2.6 for example) with its own set of values and properties; therefore, there can be multiple levels of abstractions [133]. Figure 2.6 shows point and interval based temporal abstractions where the (primitive) parameter points that hold at times T1–T6 are abstracted into higher level concepts according to a given context (the CCTG-522 context in this case). First, the values at T1 and T2 are abstracted into two abstraction points, LOW(Hb). These point abstractions are then joined into a LOW(Hb) interval abstraction, I1, by temporal interpolation. Similarly, I2–I4 are formed, and then I5 and I6. I5 and I6 are finally joined into a LOW(Hb) interval abstraction, I7, again using temporal interpolation.
The RÉSUMÉ system has been evaluated in several clinical domains such as protocol-based care, monitoring of children’s growth and managing diabetes as well as in an engineering domain for monitoring traffic control [133, 134].

2.3.2 The EON System

EON is an architecture that consists of a number of modular components and is intended to be embedded within a clinical information system [137]. EON is primarily used to offer decision support concerning various aspects of protocol-based care. It was one of the earlier attempts to create a knowledge base with an explicit domain model that could serve as input to a variety of problem solving methods (which are computational units, usually domain independent, that can be used as modular ‘building blocks’ [78]). Two such problem solving methods in EON are used for generating patient-specific treatment plans and determining the qualitative likelihood that a patient is eligible for the given protocol, both methods taking a standard clinical protocol description and relevant patient data as input [137].

Figure 2.6: Typical inputs to and outputs of RÉSUMÉ together with point and interval based abstractions [133].
All problem solving methods within EON use a common static knowledge base that is accessed via the Tzolkin database mediator [136] which acts as the conduit between the problem solving modules and the patient data stored in a database (see Figure 2.7). Tzolkin is implemented as a middleware server and consists of two modules (together with other control structures); these are the Chronus temporal data query system (represented by ‘Temporal Query’ in Figure 2.7, see Section 2.3.6.1 for more details) and the RÉSUMÉ temporal-abstraction system (represented by ‘Temporal Abstraction’ in Figure 2.7; discussed previously). When a problem solving method needs to resolve a question, such as “was there a past episode of moderate anemia that lasted for more than two weeks?”, it passes the query to Tzolkin which would first determine that anaemia is not a primitive datum stored in the patient database. It would then reason that anaemia is an abstraction of haemoglobin – a datum that is stored in the patient database. Tzolkin module would then use Chronus to query the patient database for the haemoglobin values, and then use the RÉSUMÉ subsystem to determine any sequences of low haemoglobin values that constitute intervals of moderate anaemia [137]. Another example that is relevant to the work in this thesis would be using the Tzolkin system to answer a question such as “should this patient be classified with hypertension?”; the specific problem-solving method related to classifying a patient can pass this question to the Tzolkin module which would reason that hypertension is not a primitive datum stored in the database (assuming that diagnosis details are not stored in the patient database). Tzolkin can then use Chronus to retrieve patient BPs from the database and then use the RÉSUMÉ system to determine whether there were at least three consecutively high BPs (possibly using the last three BP measurements). The results from RÉSUMÉ can subsequently be used to determine whether or not to classify the patient with hypertension.
Other than Tzolkin, EON consists of several other middleware servers that perform the computation necessary to support specific tasks in guideline-based patient care [79]. The Padda guideline execution server, for example, takes formalised clinical guidelines and relevant patient data as inputs to generate situation-specific recommendations while the WOZ server provides explanation services for other components. For performing guideline-based application tasks, EON provides a framework for structuring patient information via a patient-data information model, structuring medical concepts via a medical-specialty model, and structuring guideline knowledge via a guideline ontology/information model (called the Dharma guideline model). These models are represented on the left in Figure 2.7, and the reader is referred to [79] for further details. The EON project funding ended in March 2003; however, it is still being used within the ATHENA CDSS (see Section 2.2.1) and the SAGE project [84] which has carried forward some of EON’s work.

2.3.3 The Medical Database Adaptor Framework

The Medical Database Adaptor (MEIDA) [138] is a framework that is focussed on facilitating the use and reuse of decision support applications and knowledge bases with heterogeneous clinical databases by overcoming the heterogeneity of the clinical databases. The motivation behind this work was that the terms that one CDSS uses to describe patient information are not often recognised by another CDSS or a local clinical database. As a possible solution, MEIDA uses a virtual schema for patient data based on the HL7 RIM – a static object-oriented model in
UML notation. The framework uses this virtual schema to enable interoperability between a CDSS knowledge base and a given EMR structure. Various terminology wrappers are used to encapsulate access to local databases and create the mapping between local database structure and standard medical vocabularies, notably Logical Observation Identifiers Names and Codes (LOINC) [139], Current Procedural Terminology (CPT) [140] version 4, International Classification of Diseases (ICD) version 9 [141] and National Drug File (NDF) [142] version 4. The MEIDA framework focuses on three modes of operation: the knowledge base specification time, database setup (mapping) time and runtime data access. The knowledge base specification time is the phase where a local clinical domain expert embeds various terms and units to create a knowledge base that is grounded in standard terms and units. During the database mapping phase, each local database is mapped to (i) the MEIDA virtual medical record schema, (ii) a set of standardised terminologies and (iii) a set of standard units using the different support tools MEIDA provides to facilitate this mapping process. These mappings allow a CDSS to formulate queries at runtime using standard schema, terms and units from the knowledge base without requiring any knowledge of the underlying schema and terminology of the local database.

2.3.4 IDAN/KNAVE II Project

The primary focus of the IDAN architecture is to automatically create abstractions (i.e., short, informative and context-sensitive interpretations) of time-oriented clinical data and to be able to execute queries about such abstractions [143]. For example, myelotoxicity is an abstraction derived from the raw values of white blood cells (WBC) and platelet counts using a classification function that is defined in the context of post bone-marrow transplantation. Performing this type of temporal abstraction from raw data requires intelligent integration of the knowledge sources, data sources and abstraction services to perform the required level of context-sensitive abstraction. IDAN distinguishes between knowledge and data in that ‘knowledge’ is generic and not specific to any particular patient; for example, classification functions, clinical guidelines, clinician experiences and so on. On the other hand, ‘data’ refers to facts associated with a particular patient, such as the WBC and platelet counts stored in clinical databases. The default abstraction service for IDAN is provided by the ALMA system which uses the KBTA method (see Section 2.3.1) for solving the temporal-abstraction task using the domain’s temporal-abstraction ontology [134]. KBTA, discussed previously, is essentially a knowledge-level representation of the temporal-abstraction task which decomposes the task into parallel subtasks where each subtask is solved by a non-decomposable temporal-abstraction computational unit called a mechanism. The different components of the IDAN architecture are shown in Figure 2.8.
Figure 2.8: The components of the IDAN architecture [143]. The numbers (1, 2, ..., 8) indicate the runtime sequence of data flow when a query needs to be evaluated for a given patient.

In Figure 2.8, the top level module called the mediator coordinates the interactions between the different modules to perform the three core IDAN services: data services, knowledge services and abstraction services. According to [143], when the user of a CDSS needs to evaluate whether one of the stored queries holds for a certain patient, the application first connects to the mediator and specifies the URLs of the knowledge base and data sources that need to be applied to the patient. The application then passes the query along with the specified patient to the mediator (see Action 1 in Figure 2.8) which then retrieves the relevant knowledge-level concepts from the knowledge base (Actions 2 and 3). Any relevant primitive concepts, i.e., concepts that do not depend on other concepts such as platelet count and WBC are then retrieved from the local data source by specifying the concepts in standardised terms (4). Step (4) on its own may include four further steps: the terminology adaptor determines any local terminology that may have been used and maps the standardised terms to local terms using mapping tables (4.1), the schema adaptor then retrieves the required data from the local database using the database proprietary language, such as Structured Query Language (SQL) or Xpath (4.2), and returns the data back to the schema adaptor (4.3). If the local database units are different from the standardised units, a conversion of the units is performed (4.4) before the data is returned to the mediator (5). At this stage, the mediator has the required primitive data
and the knowledge to perform temporal abstractions, so this information, along with the original query, is sent to the abstraction service (6), where the knowledge and the query are transformed into a set of rules in the *temporal-abstraction rules* language. These rules are applied to the data by the ALMA temporal-abstraction module (6.1) which processes the rules and returns an abstract-concept answer set (6.2) back to the mediator (7) and finally to the querying application (8). The ALMA system IDAN uses is a more recent version of the RÉSUMÉ temporal abstraction system (see Section 2.3.1) and the mediator is analogous to the Tzolkin architecture discussed previously. In Tzolkin, however, the temporal-reasoning and temporal-maintenance tasks are performed by different modules – the RÉSUMÉ system module generates all abstractions and writes them to the database while the Chronus module applies a query’s temporal constraints to the database (which now also includes the abstractions) to generate the query results [143]. As well as reasoning, the ALMA system can also process a query’s temporal constraints and therefore IDAN provides a more uniform implementation than previous attempts.

An important feature of the IDAN framework is the use of standardised medical terminology [143]. The standard systems that are currently supported are: ICD-9 for diagnoses, CPT codes for procedures and LOINC for laboratory tests. NDF for drug information and the Systematised Nomenclature of Medicine – Clinical Terms (SNOMED-CT) [144] as an additional terminology standard for diagnoses (i.e., in addition to ICD-9) are also being considered [143].

KNAVE-II [145] is an intelligent interface to the IDAN framework that supports interactive exploration and visualisation of time-oriented clinical data. Similar to the ‘Decision support application’ in Figure 2.8, KNAVE-II interacts directly with IDAN to perform temporal abstraction among other tasks. The main interface of the KNAVE-II client is shown in Figure 2.9.
The different graphical panels in Figure 2.9 show the selected data at varying abstraction levels. For example, the third, fifth and sixth panels from the top show raw data types while the first, second and fourth panels show abstract concepts derived from the raw data. The temporal abstraction task is performed by IDAN in accordance with the temporal abstraction ontology as discussed in [143]. KNAVE-II is a comprehensive visualisation tool that supports a range of other advanced features, such as distributed knowledge bases, descriptive statistics and dynamic sensitivity analysis on raw and abstract data, five different zoom-in functions for visualising the temporal dimension, absolute and relative time-lines, context-sensitive explanation functions and concept searching. Refer to [145] for a detailed description of each of these features.

2.3.5 VISITORS System

The KNAVE tool discussed previously focusses on visualisation aspects of raw/abstract data as applicable to a single patient. At the same time, there is a need to analyse and visualise data in an aggregated fashion for a group of patients for tasks such as analysis of clinical trials of new drugs, quality assessment purposes as well as discovering certain patterns through the analysis of multiple patients [146]. Klimov and Shahar [146-148] propose the VISualization of Time-Oriented RecordS (VISITORS) system as a solution to address these issues and to provide
support for knowledge-based temporal reasoning for deriving temporal patterns and abstractions, information visualisation and exploration capabilities and supporting the analysis of associations among multiple patient records. The VISITORS system is essentially an enhanced extension of the KNAVE-II system and provides data aggregation and exploration features for multiple patients at various temporal granularities, such as hour, day and month (or other specific time periods). The system enables users to specify temporal and knowledge-based constraints interactively through a graphical query-builder to specify the patient subsets for exploration [148]. The query-building interface enables three types of queries: Select Patients (Who?), Select Time Intervals (When?) and Get Patient Data (What?), and each query will retrieve either a list of patients, a list of relevant time intervals or a list of time-oriented patient datasets, respectively. Various additional constraints can then be dynamically applied to the patients, time periods and data values for further analysis. Figure 2.10 shows the main VISITORS interface for a group of 58 patients that have been retrieved as a query result.

![VISITORS Main Interface](image)

Figure 2.10: The VISITORS main interface [147]

The panels denoted by ‘A’ display lists of patients and lists of time intervals returned by the queries while the graphs in panel ‘B’ show the data for the 58 patients. The top graph in panel B shows all of the WBC test values during March 95 with the red line representing the daily maximum value. The second graph shows the daily mean value of haemoglobin for each patient during 1995 where the top red line and bottom blue lines represent the monthly maxima and minima values respectively for this group of patients. The third graph shows the monthly distribution of the haemoglobin state-abstraction values (e.g., normal or moderately_low) for each
month of 1995, and the tooltip indicates how in April 1995, 73% of patients in this group had a moderately low haemoglobin value. The last graph displays the daily distribution of the platelet state-abstraction values for March 1995. The left panel is a browser for the underlying ontology that is relevant to the visualisation where the required domain ontology is retrieved on the fly from the corresponding domain knowledge base [147].

Discovering new interrelations or patterns, especially temporal interrelations within a selected group of patients is important, and to support this the VISITORS system has been augmented with a technique called Temporal Association Charts (TACs) [148]. TACs support the investigation of temporal and statistical associations within multiple patient records among raw concepts as well as the temporal abstractions derived from them. These charts enable users to graphically explore and analyse the time and value associations among domain concepts that explicitly or implicitly exist within multiple time-oriented patient records. In addition, TACs can be used to analyse the effects of constraining one parameter on the association between multiple concepts [147]. Evaluation of TACs by clinicians and medical informaticians have shown that TACs are functional and usable and can be used to answer complex time-oriented clinical statistical-aggregation questions [148].

2.3.6 Temporal Extensions to Relational Databases

Adlassnig et al. [128] identified temporal clinical databases as one of the four areas of important temporal research in medicine. Querying time-stamped data is necessary for many clinical decision support and other temporal reasoning applications, yet standardised temporal querying methods are not widely available for application developers. Most commercial database systems are based on the relational data model and the various limitations of these relational database systems in supporting the semantics of time were first noted over 15 years ago [149]. Relational database systems can readily store time values as attributes, but have an inherent limitation with respect to querying. The widely used standard SQL has very few constructs other than the simple before (via the ‘<’ operator) and after (via the ‘>’ operator) type operations to manipulate time values. This makes formulating complex temporal queries in SQL a very difficult task for users. As a result, the temporal database community has been researching techniques to extend the relational model to support temporal querying. In 1992, Richard Snodgrass proposed temporal extensions to the SQL-92 language standard which resulted in the TSQL2 Temporal Query Language Specification [149]. The more recent ISO standard version of the SQL language is SQL3 and a number of propositions were made under the SQL/Temporal project to support some of the proposals from TSQL2; however, these were
withdrawn for a number of reasons [150]. As a result, commercial database systems provide minimal temporal query support, and various research projects have evolved to bridge this gap between the temporal query requirements of the medical community and existing database query support. This section briefly outlines some of such important projects.

### 2.3.6.1 Chronus and Chronus-II

Chronus is a database mediator/query system developed at Stanford Medical Informatics that extends the standard relational model and the SQL query language to support temporal queries on relational databases. It provides an expressive SQL-based temporal query language that allows users to write complex temporal queries [151]. The researchers first analysed the requirements of the kinds of temporal data used in clinical care and the types of temporal operators needed. It was found that the temporal representation in the database should store three types of data:

- **time-stamped data** – this is the type of data that relational databases usually have support for; it is a simple, instant time stamp, such as 05-Jan-10 10.23 am.
- **interval stamped data** – these data types are used to represent durations, such as the duration of a prescription
- **temporal granularity** – this refers to the resolution at which time is recorded [152]. Some examples are days, months and years. For the work presented in this thesis, I have used a temporal granularity of one day.

To cater for these requirements, the Chronus project used a temporal database system and added the ability to support time interval semantics with a temporal dimension added at the level of the database tuple. The system was associated with a temporal extension to SQL, called TimeLine SQL (TLSQL), to support temporal querying. The underlying temporal model of the Chronus system can be used to manipulate time points and intervals with a set of temporal operations that can (1) make comparisons between time points and between intervals, (2) determine the time difference between two time points via a DURATION operator, (3) create a new time point by adding or subtracting a time point with a given granularity, and (4) compute the temporal intersection, difference and union of two adjacent or overlapping intervals. TLSQL has been applied in several clinical contexts with encouraging results, for example, in protocol-based decision support systems for HIV patients [153, 154] and in screening patients for clinical trials [151].
The temporal model and the query language of Chronus had a number of shortcomings. For instance, Chronus assumed all tables were temporal, so a Chronus database schema was incompatible with a standard relational schema and could not work with legacy non-temporal data. Also, many temporal queries had to be expressed as a sequence of two or more sub-queries requiring temporary tables to hold intermediate results. Chronus-II [152] was designed and developed to overcome these issues. Chronus-II is implemented in Java and is designed to operate as a layer above existing relational databases. It takes a temporal command, generates standard SQL statements for the non-temporal part of the command, and completes the processing of the temporal part in memory. The results are then passed to the user or written to a database. It interacts with the database through a JDBC interface and is not tied to any particular database implementation. Chronus-II is designed as a modular architecture that encapsulates the temporal database functionalities, making it easy to integrate into other projects that need temporal query support. For example, Chronus-II has been successfully used in the ATHENA CDSS (discussed in Section 2.2.1), guideline based quality assessment [152] and also in evaluating patients for clinical trials [155]. However, Chronus is no longer an active project.

2.3.6.2 TimeDB

TimeDB [156] is a freely available temporal extension to relational database systems. It supports a query language, a data manipulation language, a data definition language and assertions/constraints. Essentially, TimeDB runs as a frontend to the commercial (relational) database system such as Oracle and uses a temporal SQL dialect called ATSQL2 as its query language (for queries, updates and assertions) that gets compiled into sequences of SQL-92 statements which are then executed by the backend (i.e., the commercial database).

2.3.6.3 DXtractor

DXtractor [157, 158] is a purpose-built solution to assist clinicians at the Boston Children’s Hospital to easily retrieve patient subpopulations from the hospital database. The required functionality was conceptualised by first developing a series of prototypical questions with clinical input. The queries involve analyses that demonstrate significant temporal complexity, for example, questions such as “Generate a list of type I diabetic patients who have each had a haemoglobin A1c (HbA1c) value greater than 10 percent, excluding the first such value” [158]. DXtractor contains a GUI that can generate standard SQL queries within contexts such as ‘Doctor’, ‘Clinical’, ‘Diagnosis’ and ‘Lab’, which respectively will select a group of patients based on which doctor cares for them, the specific clinical finding of a physical exam, one or
more diagnoses, or a specific laboratory finding. Once such a patient population has been retrieved, DXtractor attempts to answer the required clinical queries with a temporal component by using a well-tested, point-based logic module that can execute temporal interval and point operations within and between retrieved patient population sets. DXtractor’s temporal operators can be combined to create complex temporal queries and the expressivity allows DXtractor to specify the 13 Allen Operators (discussed at the beginning of Section 2.3). As such, although DXtractor is perhaps not as compact as some of the other temporal query languages (Chronus-II, for example), DXtractor allows for specification of all necessary temporal functions within the selected context, without the need for database modification or SQL augmentation. The system is implemented in Java and uses Java JDBC methods for database operations, and is accessible to clinicians via any Java-enabled web browser. DXtractor has been used successfully to identify patient subpopulations of interest to clinicians, to perform simple set operators to manipulate identified groups of patients and also to perform temporal operations that are effective in defining time based relations between subpopulations [157]. However, DXtractor is no longer under active development.

2.3.7 Temporal Rule Mining

Rule mining refers to discovering interesting (and often unexpected) rules and patterns from large datasets. These patterns (or rules) may have a form such as “if we encounter element \( x \), then we will encounter element \( y \) within \( t \) time units” where \( x \) is referred to as the antecedent, and \( y \) is called the consequent. The quality of such rules may be measured by how frequently they occur (support), their predictive power (confidence) and by measures of how interesting they are, described numerically by so-called interestingness measures [159]. Temporal rule mining on the other hand refers to mining for interesting, frequent temporal associations between diagnostic or therapeutic patterns and can be used in practice to monitor the quality of care for certain pathological conditions and specific clinical conditions characterising a population of interest. Temporal association rules (TARs) [160, 161] can be used for temporal rule mining where a TAR is defined as a relationship specified through a temporal operator which holds between an antecedent consisting of a pattern of single or multiple cardinality, and a consequent consisting of a pattern with single cardinality. A pattern is defined as the occurrence of one or more contemporary events (e.g., ‘HbA1c test’ together with ‘Cholesterol test’) where the temporal relationships are specified using a fixed set of operators, such as the Allen Operators. The TAR algorithm offers the additional opportunity to select specific rule templates that define the event classes allowed for the antecedent and the consequent selection
respectively to narrow the focus of the search to relationships only between the members of the classes that the user wants to investigate [161].

TAR mining was carried out in [161] on the Regional Healthcare Agency database of Pavia where the antecedent selection was limited to state (e.g., ‘low HbA1C’) and trend (e.g., ‘increasing HbA1C’) abstractions, while the events allowed for the consequent selection were limited to drug prescriptions. This configuration was used to establish which abstractions on physiological parameters frequently showed a significant temporal association with subsequent prescriptions of drugs in order to interpret the medical prescription of the specific drugs as a reaction of the clinical conditions observed during the visit. After the rules were mined, each important rule was classified on the basis of the presence or absence of any clinical evidence of a relationship between the physiological variables included in the antecedent and the effects of the prescribed drugs represented in the consequent. An identified rule out of this study was that “a diabetic subject showing a cardiovascular risk between 5-10% and a high HbA1C value, has a 63% probability to undergo a prescription of oral hypoglycaemic agents in the following 90 days”. This rule reflects a clinical relationship between the antecedent and consequent concepts, as blood glucose lowering drugs indeed have an effect on HbA1C. In another study using the same database [160], the relationship between the antecedent and the consequent was characterised by a gap constraint, which sets the maximum allowed distance between antecedent and consequent. This study found that “a subject found to be overweight, under antihypertensive therapy, showing a normal glycaemic value and a high HbA1C has a 56% probability to have an increase in glycaemia within the following inspection (gap=1 visit)” and that “given the simultaneous prescription of high-ceiling diuretics and antigout agents, the diabetic population shows a 57% probability to have an increase in HbA1C during the following year (gap=1 year)”. Evaluation of the rule classification system based on the clinical meaning of the concepts proved to be effective and the method has been shown to be capable of characterising groups of subjects, highlighting interesting frequent temporal associations between diagnostic/therapeutic patterns and patterns related to the patients’ clinical condition of interest [160, 161].

TAR data mining techniques are suitable when a pattern of interest occurs frequently. However, these techniques are ineffective when a pattern is infrequent, as is the case with adverse drug events. Adverse drug events usually have low incidence rates (i.e., they are ‘unexpected’) as drugs are tested prior to releasing them onto the market. As a result, adverse drug events with incidence rates less than 0.1% are normally not detected and may go unnoticed until large
numbers of users have been affected. Jin et al. [162, 163] propose the use of unexpected temporal association rules (UTARs) to describe these unanticipated episodes and provide an algorithm for discovering them. UTARs are similar to TARs, with an antecedent and a consequent and a temporal relation between them. To handle the unexpectedness, an interestingness measure called \textit{residual-leverage} is used to provide a case-based exclusion technique for its calculation. The idea underpinning this technique is to exclude ‘expected’ events from a given set of sub-sequences (i.e., create a subset of temporal events that unexpectedly occur prior to the consequent) and then aggregate unexpectedness over all the remaining sub-sequences [163]. Detecting adverse drug events is of great value with regard to correction and prevention of such events, especially if outcomes are life threatening. As such, the Queensland Linked Data Set (consisting of data from several healthcare facilities in Queensland, Australia) has been used to provide a demonstration of using UTARs on de-identified administrative health data. During this effort, the investigators mined several known adverse drug events such as alendronate possibly causing esophagitis [162, 163] and atorvastatin possibly causing stomach ulcers [162] as well as some previously unknown adverse drug events, for instance, flucloxacillin possibly causing hepatitis [163]. A new tool called the \textit{iHealth Explorer} with a web interface is being developed as a web service to deliver UTAR type data mining and analytic facilities in an attempt to provide desktop access to very large data collections [164].

2.4 Quality Indicators in Primary Care

The previous sections discussed CIG modelling techniques, guideline based decision support systems as relevant for chronic disease management and various other projects that have evolved around guideline based treatment (which usually requires the use of temporal analysis and visualisation tools). CPGs present the best available scientific evidence reviewed and adopted by professional organisations. Therefore, along with the ‘at the point of care’ support systems, it is also important to monitor/audit adherence to these guidelines using quality indicators which are derived from the evidence in CPGs.

Drawing attention to applying current, professional knowledge in treatment and diagnostic processes in healthcare, the Institute of Medicine defines quality of care as “the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge” [165]. Campbell et al. [166] argue that such definitions are very generic, and not easily operationalisable as they trade both sensitivity and specificity for generalisability, and propose a more refined definition where
quality of care is “whether individuals can access the health structures and processes of care which they need and whether the care received is effective.” Quality of care is a multidimensional concept and different methods of measurement need to be considered for different aspects of quality [167, 168]. Despite a range of definitions and the lack of a universally accepted definition, for the purposes of this thesis, quality of care is considered to be measures of outcome, a definition used by other authors as well [167, 169]. Similarly, quality indicators are primarily used to give an indication of the quality of the patient care delivered and have been devised to measure and document performance to motivate clinicians (and hence their organisations) to improve through the use of a common metrics [4].

Three broad types of indicators can be found in the literature: (a) indicators of the structure of a healthcare system, (b) indicators of the processes of care and (c) indicators of the outcomes resulting from care [4, 166]. Structure refers to organisational factors/conditions that define the health system under which care is delivered, for example, the ‘percentage of teams for diabetes care including a foot therapist’ or ‘existence of a stroke unit’ [4]. Processes of care usually involve the various interactions between users and the healthcare structure, can be influenced directly and measured how (often) something is done [4, 166]; in essence, this is what is done to or with the users and may include indicators such as the ‘percentage of diabetic patients getting an annual eye test’ [4]. Outcome measures on the other hand give information on the outcome of care processes measured at patient level, the ‘percentage of patients with severe pain at 36 hours after surgery’ for example [4]. Outcomes may be influenced by structure as well as processes, indirectly or directly. For example, a patient may die from cervical cancer either because a screening service was not available (structure) or because the cytology report was misread (process) [166].

Development of quality indicators is a nontrivial task. The many important dimensions and issues involved in the development process of these indicators are discussed in [170-172]. It is imperative for quality indicators to be meaningful, scientifically sound, generalisable and interpretable, and therefore they must be developed, tested and implemented with scientific rigour [4]. Mainz [173] describes the steps involved in developing and testing clinical indicators with respect to the Danish National Indicator Project [174], where it is suggested that the development process should include a planning phase (to evaluate the clinical areas of interest and have the measurement team organised) followed by a development phase where quality indicators are prioritised and selected based on scientific evidence and literature. The specific measures should then be designed to include inclusion and exclusion criteria for the target
population, a description of a risk adjustment strategy, identification of data sources, a description of data collection procedures and an analytical plan for data analysis. Finally, the clinical indicators should be tested for reliability and validity before they are implemented as preliminary tests may identify areas requiring further modifications and specifications of the indicators. However, it should be noted that these quality indicators are not intended to set optimal standards of care for any individual patient. Any resistance to following evidence based guidelines or other conflicts between CPGs and quality indicators that may arise during the development and implementation phases should not mean that these should be abandoned, but rather that they need to be refined and used appropriately to achieve their ultimate goals – i.e., promoting quality (through changes in practice and/or selection) and ensuring that medical care is based on scientific evidence [175].

Over the last few decades, researchers have investigated various aspects of quality audit frameworks for primary care and proposed a number of different schemes. Most of the frameworks have been evidence-based theoretical/conceptual frameworks that have been developed using surveys, consensus and domain expertise [166, 176-182]. Some studies have focussed on quality indicators targeting specific problem domains (such as hypertension [183], diabetes [184], cardiovascular disease [185] and mental health [186]) while a few studies have taken a much broader approach and attempted to develop or have proposed multi-national level quality indicators. As an example of the latter, consider the European Practice Assessment instrument and indicators that can be shared by six European countries (Belgium, France, Germany, Netherlands, Switzerland and the UK) [187] and has been subsequently validated via a feasibility study using 273 general practices in Austria, Belgium, France, Germany, Israel, Netherlands, Slovenia, Switzerland and the UK [188]. Another large scale project that is focussed on developing quality indicators across various healthcare domains is the OECD Health Care Quality Indicators Project [179, 189], which also attempts to develop quality indicators for primary care across 21 different countries [190].

Several national level attempts have been made to develop quality indicators for primary care, most notably from Canada, Denmark, NZ and the UK. The Pan-Canadian Primary Health Care Indicator Development Project [191] identifies 105 primary care healthcare indicators across eight different categories, agreed upon by a broad audience of stakeholders from across the country. An EMR dataset to facilitate this quality audit effort has also been proposed; this includes data elements related to patients, providers, encounters and outcomes of care [192]. The Danish National Indicator Project [174] was established in 2000 and implemented in 2003
with the intention of documenting and developing the quality of care for the benefit of patients. Since then, evidence based specific quality indicators have been developed for a wide range of medical conditions, such as diabetes and heart failure. A condition usually has 6-10 indicators that reflect the medical technical aspects of care in terms of prevention, diagnostics, treatment, care and rehabilitation. NZ has a well established Primary Health Organisation (PHO) Performance Management Programme [193] for primary care where a range of performance indicators have been developed by primary care representatives, District Health Boards (DHBs) and the Ministry of Health in an attempt to improve the health of enrolled populations and reduce inequalities in health outcomes\(^1\). PHOs receive incentive payments for improvements in performance that are measured against a range of nationally consistent indicators. Standards have been established for the operational framework [194], reporting/data formats as well as the content of PHO clinical performance indicators [195]. The Quality and Outcomes Framework (QOF) in the UK [196] is perhaps the world’s single largest attempt to improve the quality of primary care wherein “with one mighty leap, the NHS vault[ed] over anything being attempted in the United States, the previous leader in quality improvement initiatives” [197]. The QOF was introduced in 2004 as part of the General Medical Services Contract and cost over 2.5 billion pounds in its first two years of operation [198]. The QOF is essentially a set of clinical indicators across four domains – clinical, organisational, additional services and patient experience [196] that have been designed around best practice with the intention of improving the quality of service provided to patients. Each indicator is allocated a number of points and GP practices are awarded points according to how well the practice has performed, and subsequently paid according to how many points they have scored. The indicators are usually updated on an annual basis.

Australia and the US have well established organisations that are actively developing healthcare quality indicators. However, these two countries currently do not have a nationally based performance measurement scheme. All States in Australia have established their own performance measurement schemes, although the Australian National Health Performance Committee and the Australian National Health and Medical Research Council are currently working towards establishing a National Health Performance Measurement Framework. This framework consists of three performance tiers – health status and outcomes, determinants of

\(^{1}\) NZ has a fairly hierarchical structure for healthcare management; the NZ Ministry of Health provides public funding to the 21 DHBs in the country which are responsible for the ongoing management of the PHOs in their areas; all NZ citizens and permanent residents can be enrolled with one PHO, which is funded for management of that person; each general practice is associated with a PHO. All analyses in this thesis are for enrolled and funded patients only, unless specified otherwise.
health and health system performance – that also includes indicators for primary care [199, 200]. Operating across State boundaries, the Bettering the Evaluation and Care of Health (BEACH) program [201, 202] in Australia is an initiative conducted by the Australian General Practice Statistics and Classification Centre to provide information (as a ‘snapshot’ in time) about the content of GP-patient encounters, the problems managed and the treatments provided by GPs to the Australian community. It is the only continuous national randomised study of general practice activity in the world and relies on the cooperation of randomly selected GPs across Australia where about 1000 GPs participate each year [202]. Each GP completes details for 100 consecutive patient encounters on structured encounter forms and provides information about themselves and their practice [199, 200]. Other than these state-wide initiatives, the tools developed by Pen Computer Systems in Australia [203] are of particular interest for this thesis. Pen Computer Systems is a commercial vendor that has developed a widely used reporting tool called the Clinical Audit Tool (CAT). The CAT integrates with existing PMS systems to provide population reporting enhancement; for example, to determine ‘diabetic patients who have not had an HbA1c test in the last 12 months?’ or ‘diabetic patients with an HbA1c >7%?’ CAT provides a population view of practice data to identify gaps in therapy and shows the practice exactly where these gaps exist so that action can be taken accordingly.

The US does not have a single national healthcare performance measurement system, but the National Committee on Quality Assurance has developed a performance measure matrix called the Health Effectiveness Data and Information Set (HEDIS) which is used by 90% of US health plans to measure managed care performance [204, 205]. Several other local/organisational level quality measurement projects exist, for example, the Physician Quality Reporting Initiative project by the Centers for Medicare and Medicaid [206] and the California State Pay for Performance Program [207].

The advantages of using quality indicators for tasks such as providing feedback for quality improvement initiatives, accreditation and regulation are well documented [208, 209]. However, associating these measures with financial incentives (i.e., ‘pay for performance’ schemes) has met with considerable debate to-date [210-212]. The intended outcomes of such a scheme are explicit, yet some investigators have characterised the unintended consequences of such schemes as worrisome, unknown and, in many instances, immeasurable [211]. Aspects such as scheme objectives, units of assessment, performance measurement analysis and interpretation of performance data, performance standards, financial rewards, current knowledge and getting benefits without problems should be considered prior to designing such a scheme to minimise
unintended consequences [210]. These may not necessarily be predictable [213], for example, there have been cases in the UK where patients were kept in hospital corridors and sometimes even in ambulances parked in car parks and holding bays before the patients could join the queue for urgent treatment so that hospitals could “meet Government targets to treat people within fours hours of admitting them” [214]. On the other hand, it is required for ambulance services to respond to 75% of calls of immediately life threatening emergencies within 8 minutes, and as a result, there have been cases where untrained staff and volunteers have been used to act as ‘community first responders’ to get to emergencies ahead of paramedics [215]. To overcome such unintended consequences, Bevan and Hamblin [216] propose that measures to counter ‘gaming’ should be included as an integral part of any audit system.

Being the largest pay-for-performance scheme, most of the expressed views and studies that have evaluated the impact of pay-for-performance schemes have been QOF related [211, 212, 217-219]. The QOF has met with criticism from various investigators, for instance, Mangin and Toop [211] argue that the QOF by its nature promotes simplicity over complexity and measurability over meaningfulness and that the QOF rewards GPs for what they have already achieved. They suggest that the impression that GPs performed much better than expected after the QOF was implemented has reinforced the widely held perception that British GPs will only do something worthwhile for additional money, which threatens public confidence in and respect for the profession. Fleetcroft et al. [220] indicate that excessive exception reporting (a provision to exclude patients from the treatment indicated in a given indicator) and having maximum threshold targets less than 100% (e.g., the minimum threshold to receive the maximum incentive for CHD-10 criterion is 50%) in the QOF introduce an incentive ceiling, which substantially reduces the percentage of eligible patients that UK practices need to treat in order to receive maximum incentive payments for delivering that care. Several studies have reported that the impact of the QOF on care is not straightforward in situations like emergency admissions and mortality [218] and the management of diabetes [219] and coronary artery disease [221]. Despite some of this scepticism, the use of QOF indicators has been associated with positive outcomes, such as improvements in BP monitoring and control [222] and better and more equitable management of coronary heart disease across ethnic groups [223], as well as managing diabetes [224] and asthma [221]. The QOF has been a bold initiative by the UK; developing quality indicators for performance measurement efforts (irrespective of whether they are linked to financial incentives) is an ongoing effort, and countries such as the US and NZ are drawing upon lessons learnt from the QOF in an attempt to enhance their current primary care healthcare systems and processes [182, 204, 213].
Developing primary care quality indicators is difficult because of the effort required, but the actual implementation and integration of these quality indicators into the already-busy schedule of clinicians is perhaps harder. Providing active decision support at the point of care is important, as is providing feedback to clinicians via successful quality audit programs and maintaining this feedback loop to ultimately improve outcomes for patients. However, research has reported a number of barriers to implementing such processes successfully, with organisational issues being a key barrier. Following the ‘Learning Organisation’ rationale suggested by Peter Senge [225] may be a potential solution to addressing such barriers. Senge defines Learning Organisations as “organizations where people continually expand their capacity to create the results they truly desire, where new and expansive patterns of thinking are nurtured, where collective aspiration is set free, and where people are continually learning to see the whole together”. In a primary healthcare environment clinicians want to treat patients vigilantly to improve their health outcomes; CDSS developers want clinicians to actively use the CDSSs they develop to support and enhance the diagnosis and treatment process; CPG developers want to promote the latest clinical findings and expect clinicians to integrate the latest recommendations into practice; policy developers attempt to develop policies/guidelines taking fairness, legal issues, resource utilisation and other factors into account; and quality audit indicator developers develop indicators that they anticipate will enhance feedback to clinical practice. As such, there is a myriad of relationships that need to be fostered for a healthcare setting to be successful and achieve its goals.

A Learning Organisation has five main features – systems thinking, personal mastery, mental models, shared vision and team learning. With respect to a healthcare setting, if the people involved can be given a better appreciation of the systems involved (i.e., getting them to look at the whole system rather than focussing on the individual parts and seeing the organisation as a dynamic process), it will lead to better outcomes. Personal mastery is important, as people with a high level of personal mastery are acutely aware of their ignorance, their incompetence and their growth areas, and they are deeply self-confident. Mental models within an organisation involve seeking to distribute activities responsibly far more widely while retaining coordination and control. A shared vision is important, as it has the power to be uplifting and to encourage experimentation and innovation. The last feature is perhaps the most important, where people within the organisation need to be able to act together. When teams learn together, Senge suggests that not only can there be good results for the organisation, but it will also help members to grow more rapidly than they could have otherwise. All people have the capacity to
learn, however the structures in which they must function are often not conducive enough to reflection and engagement. Paying sufficient attention to the details of a Learning Organisation and working towards developing the suggested attributes may result in better coordination, collaboration and an understanding between various groups involved in healthcare, all with similar ultimate goals.
Chapter 3

Development of Quality Audit Criteria for Chronic Disease Management

The purpose of this chapter is to provide an overview of the quality audit criteria development process as relevant to this thesis and present the criteria that have been developed. The chapter first provides some background information related to some initial work that laid the foundation for the work presented in the rest of the thesis. This is followed by a discussion on the criteria development process and details of a validation exercise that was carried out to ensure the validity of the audit criteria. The chapter concludes with a discussion on several high level categories of audit criteria that have been identified to provide the maximum reasonable level of abstraction from the developed criteria.

3.1 Background – the Therapeutic State Transition Model

I started my PhD research around the time a series of meetings were being held (refer to Section 3.2 for details) between The University of Auckland based researchers and an expert panel from a general practice the researchers were collaborating with. The requirements of the expert panel included quality indicators such as ‘patients who have been on concurrent therapy with angiotensin-converting enzyme inhibitors (ACEi)/angiotensin receptor blockers (ARBs) and diuretics’ and ‘patients with lapse in antihypertensive medication’, where a lapse in antihypertensive therapy is said to commence when all antihypertensive medications, if taken
as directed from the day of prescribing, run out (brief lapses are expected and may not be problematic where the patient has retained some prior supply). I was involved with the technical implementation of these criteria, and to satisfy the expert panel audit requirements such as medication lapses, I used the therapeutic state transition model set forth by my PhD supervisor, Prof. Jim Warren and others [12] and then further enhanced by Gadzhanova et al. [10]. Within the therapeutic state transition model, CPG recommendations are first abstracted into key therapeutic states. EMR prescribing data is then mapped into the corresponding states which are then used to identify cases on suboptimal care by interpreting the transitions in patient states with respect to the CPG recommendations [12]. In essence, this model looks at analysing chronic disease prescribing actions over time in terms of transitions in status of therapy as indicated in the EMR prescribing records [10].

According to the therapeutic state transition model, each prescription produces two events - one marking the start of the prescription and another implicit event marking the expected end of the prescription if directions (dose, frequency and repeats/refills for example) given by the GP were properly adhered to. By using various state variables to denote antihypertensive therapy (this was the primary domain of interest – see Section 3.2) as shown in Table 3.1, a series of therapeutic state transitions were constructed which are effectively points in time when the status of key aspects of a patient’s therapy changes (for example, being on ACEi/ARBs – i.e., state ‘A’ in Table 3.1 and running out of medication; i.e., moving into a Zero state). This model is further enhanced by heuristically processing the states to avoid over-sensitivity with lapses (indicated by Zero state) of less than 90 days being coalesced into the prior state, as is any other state of less than 30 days duration. My implementation of the therapeutic state transition model as relevant to this thesis is described in [21].

<table>
<thead>
<tr>
<th>State Variable</th>
<th>ATC Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A ACEi and ARBs (ATC: C02EA)</td>
<td></td>
</tr>
<tr>
<td>B1 Beta-blockers (ATC: C07, especially discerning selective beta-blockers: C07AB)</td>
<td></td>
</tr>
<tr>
<td>B2 Diuretics (ATC: C03AA – thiazides and C03C – loop diuretics)</td>
<td></td>
</tr>
<tr>
<td>B3 Non-dihydropyridine calcium channel blockers (ATC: C08CX, C08DA01, C08DB01)</td>
<td></td>
</tr>
<tr>
<td>C Dihydropyridine Calcium Channel Blockers (ATC: C08CA)</td>
<td></td>
</tr>
<tr>
<td>D Alpha blockers, hydralazines and clonidine (ATC: C02DB, C02CA, C02AC)</td>
<td></td>
</tr>
</tbody>
</table>

Also included in the analysis are combination products that represent more than one of the above groups – ATC: C07BA, C07BB, C03AB, C08GA01, C09BA, C09BB and C09DA. Based on the therapeutic states in Table 3.1, a state transition overview diagram (Figure 3.1) can then be constructed for the practice, and this process is described in [10, 12]. This diagram presents a
patient count of various state transition movements within the particular practice during the defined time period.

Figure 3.1: Therapeutic state-transitions for antihypertensive prescribing over a 20-month period for patients with hypertension and diabetes (for transitions with ≥3 occurrences) [12]

The arcs in the state transition overview diagram indicate the number of patients within the practice with that particular transition. This view of the practice does not show how an individual patient has moved through different transitions, and to cater for this requirement, individual path diagrams have been proposed [11].

For a hypothetical patient, Table 3.2 shows the state variables for the prescriptions the patient received during a selected time period together with a brief explanation on the state transitions that occur. Figure 3.2 shows the corresponding individual path diagram for this patient.
Table 3.2: Prescription durations, their therapeutic state variables and corresponding state transitions for a hypothetical patient [21]

<table>
<thead>
<tr>
<th>Prescription Date</th>
<th>State variables for prescriptions and durations</th>
<th>State Transition</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>09-Sep-06</td>
<td>A (90) B2 (90)</td>
<td></td>
<td>New prescriptions issued before previous ones fully expired. No change in states since patient still has A B1 and B2 drugs. A and B2 are scheduled to expire on 08-Dec-06.</td>
</tr>
<tr>
<td>16-Oct-06</td>
<td>B1 (90)</td>
<td></td>
<td>B1 would have expired on the 30-Sep-06, but since duration between this and 16-Oct-06 is less than 30 days no state transition occurs (heuristics). B1 is scheduled to expire on 14-Jan-07.</td>
</tr>
<tr>
<td>08-Dec-06</td>
<td>A B1 B2 → B1</td>
<td></td>
<td>A and B2 expire. Patient is only on B1. Class A medications are compellingly indicated for diabetes, so an alert will need to be raised if the patient was diabetic, unless current BP measurements are in the normal range.</td>
</tr>
<tr>
<td>14-Jan-07</td>
<td>B1 → Zero</td>
<td></td>
<td>B1 expires. Patient is on no medication and alert is required, especially if no controlled BP measurements (checking of BP is required since the patient could have been taken off medication and put on diatery management).</td>
</tr>
<tr>
<td>02-Mar-07</td>
<td>A (90) B1 (90) B2 (90)</td>
<td></td>
<td>New prescriptions issued before previous fully expired. No change to states since patient is still on A B1 and B2. All prescriptions are scheduled to expire on 31-May-07.</td>
</tr>
<tr>
<td>31-May-07</td>
<td>A B1 B2 → Zero</td>
<td></td>
<td>A B1 and B2 expire. Again the patient is on no medication and alerting may be necessary.</td>
</tr>
<tr>
<td>13-Jul-07</td>
<td>C (90)</td>
<td>Zero → C</td>
<td>New prescription issued for C. Due to expire on 11-Oct-07.</td>
</tr>
</tbody>
</table>

Figure 3.2: An individual path diagram – The text on the arrow shows the transition date, the duration in each state and the total duration from the beginning of analysis period (within brackets). All durations are in days [21].

Table 3.2 is self-explanatory with the ‘Explanation’ column detailing the important observations pertaining to the model. Figure 3.2 shows a visual representation of how this patient has been in and out of antihypertensive medication. The significance of the therapeutic state transition model is that it gives the analyser a suitable means of identifying patients on concurrent therapies as well as patients with lapses. It provides a higher level of abstraction compared to extracting the information from a raw prescriptions table. This is by no means to say that the
state transition model does not have its own limitations, and these are addressed later in Section 3.3.

3.2 Development of Audit Criteria and The Quality Audit Report

Around the time I started my PhD, my supervisor was putting together a University based research team to collaborate with a general practice in West Auckland with a predominance of Pacific Island patients. I became a member of this team which conducted a series of meetings with the collaborating general practice. The practice formulated an expert panel for this work consisting of five members; two GPs, the practice manager and two nursing staff. During the first meeting the opportunity was given to the expert panel to propose suitable criteria for auditing their management of patients with chronic disease, focusing mainly on hypertension and its common comorbidities. This was the background to the audit criteria development initiative which will be described in some detail in the remainder of this section. This research was approved under the University of Auckland Human Participants Ethics Committee protocol number 2007/078.

After the initial meeting, three one-hour meetings with the expert panel were conducted on the premises of the practice between May 2007 and July 2007 to develop a quality audit report based on EMR data from the practice. The information needs for audit were determined primarily by the practice’s panel, while the University based researchers also provided their input (the University based research team had two medical doctors).

The quality audit report was designed to document:

- Descriptive data about the practice (e.g., prevalence of hypertension);
- Positive attributes (numbers that the panel would like to raise – e.g., percent of patients diagnosed with hypertension with BP now controlled);
- Quality improvement opportunities / cautionary attributes (cases registering in this category are recommended for review).

After each of the three meetings, queries were implemented to populate the quality audit report from practice data, based on data extracted from the practice’s commercial PMS (MedTech32 [226]). The protocol was to extract prescribing and relevant laboratory data and observations for the 18-month period preceding the first meeting, with the exception of problem diagnoses (based on Read Codes [227]) which are relevant for an indefinite time with respect to chronic
disease and hence were extracted for as far back as five-years. Use of unique practice identification codes for each patient (internally, MedTech32 refers to this identifier as ‘MMID’ which will be used throughout this thesis to indicate a unique patient identifier) allowed the practice to identify patients while maintaining patient anonymity in data used by the external investigators. Note that this MMID was not the NZ National Health Index (NHI) number (which is a unique patient identifier used within the NZ health system for clinical and administrative purposes), but a practice and PMS specific identifier. Also, within MedTech32, diagnoses are referred to as ‘classifications’, and therefore these two terms are used interchangeably in this thesis.

Although the extracted data span was 18-months (with the exception of classifications which was for five-years), the reporting window (i.e., the evaluation period) was narrower than this as shown in Figure 3.3. For this particular report, an evaluation period of 12-months was suggested by the expert panel. The reason for extracting prescriptions and other relevant data for 18-months (with the exception of classifications) was to develop the gradual build-up of therapy using therapeutic state transitions. This was required due to the nature of chronic disease which usually spans over an indefinite period of time, hence will be inaccurate to say that a hypertensive patient was not on antihypertensive therapy just because there was no antihypertensive prescription at the very beginning of the reporting period (as therapy that started sometime before the beginning of the evaluation period could have continued into the evaluation period). This period (six-months in this case) prior to the beginning of the evaluation period is referred to as the run-in period from here on.

![Figure 3.3: The evaluation period specific to the quality audit report](image)

Relevant laboratory tests were identified in advance of the first meeting by analysis of guidelines (notably, JNC7), but extended based on meeting results (in particular, to include estimated glomerular filtration rate; eGFR). The queries required some pre-processing to compute the duration of medication supply for each prescription as indicated by dose, frequency, pack size and repeats (generally 90-days). Lapses were identified where a
medication, if first dispensed on the day of prescribing and subsequently taken as directed, would have run out. Periods of lapse were computed both for overall antihypertensives and for several antihypertensive drug groupings of interest to the panel, including ACEi and ARB (collectively), beta-blockers and thiazide diuretics. The computation of lapse periods was based on the therapeutic state transition model discussed previously (Section 3.1) with appropriate settings for the heuristics (e.g., lapses less than 90 days were not coalesced into the prior state as shorter lapses needed to be identified).

The audit criteria developed during the meetings with the expert panel were put together in the form of a quality audit report containing three main sections – a descriptive section (with 7 criteria) with general information about the practice, a section with supportive indicators (with 12 criteria) which contains information the practice in general would want to drive to a 100 percent and a final section (with 17 criteria) which contains cautionary indicators. Table 3.3 shows two examples from each section of the quality audit report, one indicating a summative or an outcome related indicator and the other relating to a process or a treatment.

<table>
<thead>
<tr>
<th>Type of statistic</th>
<th>Summative / Outcome</th>
<th>Treatment / Process</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive</strong></td>
<td>Number (percentage) of active patients classified with hypertension as of 8 May 2007</td>
<td>Number (percentage) of BP measurements measured by GPs</td>
</tr>
<tr>
<td><strong>Supportive</strong></td>
<td>Number (percentage) of patients classified with hypertension with systolic/diastolic BP less than or equal to 140/90mmHg (on last BP reading)</td>
<td>Number (percentage) of patients with active prescriptions for more than one antihypertensive agent as of 8 May 2007</td>
</tr>
<tr>
<td><strong>Cautionary</strong></td>
<td>Number (percentage) of patients with three or more consecutive BP measurements over 160/100mmHg (last measurement during the reporting period)</td>
<td>Number (percentage) of patients classified with hypertension and diabetes with lapses in ACEi or ARB during the reporting period</td>
</tr>
</tbody>
</table>

The supportive and the cautionary sections of the quality audit report contain quality indicators under several subcategories. The supportive section constitutes of: BP related, prescription related, effective combination therapy, continuity of therapy and drug-problem indications; the cautionary section has the subcategories: generic (for indicators such as ‘patients classified with hypertension and not prescribed with antihypertensive agents within 90 days of classification’), BP related, lack of continuity of therapy, drug-drug interactions, drug-problem interactions and monitoring. An instance of an actual quality audit report that was generated with the criteria together with resulting patient numbers for this particular practice can be found in Appendix I.
Some important temporal issues that need to be considered when formulating the quality indicators are discussed in Chapter 5 where temporal requirements have been identified methodically.

3.3 An Initial Implementation of the Audit Criteria

3.3.1 Data Extraction, Data Cleaning and Pre-Processing

The first phase of the implementation of the developed criteria was visiting the practice and extracting the required data from the practice’s commercial PMS (i.e., from MedTech32) to a password-protected Microsoft (MS) Access research database. MedTech32 is widely used by NZ GPs to manage their patients; however, being a commercial PMS system, MedTech32 does not provide direct access to its internal database structure. Although the underlying database tables cannot be accessed directly, the system provides a database querying interface via a visual query builder tool that lets an interested user extract certain database fields (that includes everything we needed for our purposes) and save this information into plain text or comma-separated values (CSV) files. The query builder tool was used to extract the required data, however, a disadvantage was that if the database fields were of type *memo* (used mainly for lengthy fields such as clinical notes, and laboratory test results), the resultant files would include many control characters often used for formatting. Therefore, after the data was extracted, a data cleaning phase was required to filter out the control characters and extract the required values. Visual Basic for Applications (VBA) modules are part of the MS Access database management system suite with direct access to the data tables in the database, and therefore cleaning was performed using VBA modules.

BP measurements also needed some cleaning and pre-processing. Within the PMS, there are two ways of entering BP values so that they get stored in the database table that maintains BPs; one is to enter the values directly into the BP fields provided in the software as an observation (of type ‘BP’), and the other is to enter the values into the physician notes field as plain text with a backslash followed by the letters *bp* or *BP* (i.e., \bp or \BP). However, there were many instances where BP values were entered into the notes field without following this convention; for example, “BP - rt arm sitting 190/90” and “rpt Bp 150/85 at 2.15 pm, felt better…” . All such instances of BP measurements were extracted from the notes using VBA regular expressions and added to the BP table in the research database. Further, some BPs were correctly entered into the field provided within the PMS as well as entered into the notes field; in such cases only one record was included.
After the data was cleaned and prepared to be processed for analysis, the prescription durations were determined using the PMS’s computed duration field if it was available. If the duration field was not set, the duration was computed based on instructions (signatura) given by the GP consisting of dose, frequency, number of repeats (refills) and the quantity. In NZ, a prescription (usually issued with two refills) is valid for a period of three months before it expires (medication may be dispensed as 90-days supply at once, or may be collected one month at a time from a single 90-day prescription). However, during visual inspections of the data several anomalies related to ways of recording prescription durations were identified (see Figure 3.4).

![Table](image)

**Figure 3.4:** Some of the different ways of recording prescription durations within the PMS – (i) PER=90, RPT=2, QTY=90; (ii) PER=30, RPT=0/2, QTY=30; (iii) PER=90, RPT=0/2, QTY=30. PER refers to the PMS computed period/duration, RPT is the number of repeats and QTY is the quantity. MMID is the unique patient identifier mentioned in Section 3.2.

The usual method of computing prescription duration is \((1+\text{RPT}) \times \text{(QTY/Consumption)}\) where consumption is determined based on the SIGS (i.e., the instructions). So for example, if the SIGS
is “1 tabs, Once Daily”, consumption would be 1, and if QTY=90 and RPT=2, then duration would be 270. However, in NZ a prescription duration cannot be 270 days. Therefore cases in Figure 3.4 (i) need to be interpreted as cases where GP intention was to prescribe a total of 90 tablets over 90 days as indicated in the PER field and computing duration based on repeats, quantity and consumption will be incorrect. The reason for this inconsistency is because the PMS system usually automatically calculates the PER field based on other fields (repeats, quantity and built-in SIGS), but when PER values are entered manually/overridden (e.g., when non built-in SIGS are specified), this field may not get automatically populated/updated. Cases (ii) and (iii) in Figure 3.4 show cases where RPT=0/2 and QTY=30, however, the PER field is set to 30 and 90 respectively. Due to such anomalies, heuristics were used to populate a new ‘Duration’ field in the database instead of using PER as the duration of a prescription. This Duration field was set to 90 if PER was 90 (irrespective of RPT/QTY) or else it was computed using quantity, repeats and consumption.

The nature of chronic illness is such that once a patient is diagnosed with a chronic condition such as hypertension or diabetes, it becomes a life-long condition. Usually it is expected that patients get classified with a chronic condition within the PMS only once, and although it is indeed the case with majority of the patients, non-scientific analyses based on ad hoc SQL queries show that approximately 40% of all hypertensive patients get classified with hypertension more than once. In fact, in one extreme case, a particular patient in the patient cohort was classified with Hypertensive disease (based on Read Clinical Codes) a surprising 21 times within the space of only five years (Figure 3.5).
It is unclear why such entries for a single chronic condition for the same patient get entered into the PMS multiple times by the same physician, but it could be due to a range of factors including physicians not going through the patient history to look at any previous/existing classifications (possibly due to consultation time constraints). Whatever the reason may be, it is an existing issue and when such cases were encountered, the earliest classification date was used as the classification date for that patient for that particular condition simply due to the nature of chronic illness.

The raw data extracted from the practice were imported into the database with table names having the format `raw_Type`, where `Type` refers to prescriptions, classifications, BPs, demographics and labs. After the data was imported into the database and cleaned accordingly, several new intermediate database tables were created from the raw tables to facilitate the querying process (which was essentially the data preparation stage). These tables, such as those with updated prescription durations were named `im_Type`. Several other tables were also created to store the required knowledge for queries; notably:

- separate tables to hold classification information: `ReadCodes_HT` (to hold the hypertension related Read Codes), `ReadCodes_Diab` (to hold the diabetes related Read Codes) and so on depending on the classification requirements of the quality audit report.
- a table to hold the different drug-drug class information based on therapeutic state variables (see Table 3.1 for the different drug classes).
The important steps related to the data pre-processing stage are shown in Figure 3.6.

In order to simplify the complexity of the final queries required to generate the audit report, several intermediate data tables were created as shown in Figure 3.6; e.g., separate tables were created for patients with a given classification (e.g., all patients with hypertension would be in QAR_Classifications_HT table). Likewise, a final prescriptions table called QAR_Prescriptions_AHT was created for antihypertensive prescriptions with an additional field for the therapeutic state variable depending on the drug class of the prescription (e.g., a prescription for quinapril which is an ACEi, will contain ‘A’ in its state variable column). A QAR_NextState table was also created (after several intermediate steps implementing the
required state transition algorithms in VBA) using the QAR_Prescriptions_AHT table to indicate the various state transitions that have occurred in a patient’s therapy.

### 3.3.2 Quality Audit Report Generation

There were several options for generating the quality audit report. Using the reporting functionality of MS Access was appealing due to the low development effort required. Developing a criteria reporting engine using the reporting features of a high level programming language such as .NET (C# or Visual Basic) or Java was another option. However, after some discussion among the researchers, it was decided to use MS Word with VBA as the implementation strategy, with VBA being used for database connectivity, executing queries and other functionality. A main reason for this was the ease of creating the required report layout using an MS Word front-end; which was also the main drawback encountered while trying to use MS Access reporting features.

The quality indicators belonging to different sections (descriptive, supportive and cautionary) and subsections within them (*effective combination therapy* and *drug-problem interactions* for example) were created in MS Word using standard tables. For each indicator, a corresponding MS Access query was written in SQL – the MS Access version of it rather, which is a subset of ANSI SQL-92 – to get the required statistic for the indicator. These queries were driven using VBA modules within MS Word and the resulting values were written into the corresponding cell in the table against the indicator (see Appendix I). An overview of this process is shown in Figure 3.7.

![Figure 3.7: Overview of the QAR generation process](image)

The relevant queries were formulated while taking important temporal issues into account (these are discussed in more detail in Chapter 5). For example, to determine patients with hypertension and diabetes who have lapses in their ACEi/ARB medication at any time during
the evaluation period, the three cases in Figure 3.8 need to be taken into account since a lapse can occur according to any (or as a combination) of these base cases. Diagnosis information is not shown in the figure; however, they also need to be considered as relevant.

![Figure 3.8: Base cases to identify a lapse in medication during the evaluation period](image)

The resultant SQL query shown below is indicative of the nature of the complex SQL that is involved with most of the criteria requiring temporal constraints.

```sql
SELECT DISTINCT QAR_Classifications_HT.MMID
FROM QAR_EvaluationPeriod, QAR_Classifications_HT, QAR_Classifications_Diab, QAR_NextState
WHERE
QAR_Classifications_HT.MMID = QAR_Classifications_Diab.MMID AND
QAR_Classifications_Diab.MMID = QAR_NextState.MMID AND
QAR_NextState.Event_date >= QAR_Classifications_HT.CLASSDATE AND
QAR_NextState.Event_date >= QAR_Classifications_Diab.CLASSDATE AND
(QAR_NextState.From_state Not Like "*A*" AND QAR_NextState.To_state Like "*A*")
OR
(QAR_NextState.From_state Like "*A*" AND QAR_NextState.To_state Not Like "*A*")
OR
(QAR_NextState.To_state Not Like "*A*", QAR_NextState.Event_date <= QAR_EvaluationPeriod.Start AND
(QAR_NextState.Event_date + QAR_NextState.duration) >= QAR_EvaluationPeriod.End)
```

Hypertension and diabetes classification details are in tables QAR_Classifications_HT and QAR_Classifications_Diab respectively. QAR_EvaluationPeriod table contains information about the evaluation period and the QAR_NextState table contains the state transitions for the patient. The state transitions are recorded in the QAR_NextState table such that the From_state and the To_state collectively denote a transition in therapy as per the state transition model. The SQL query shown above is self explanatory with the three cases in Figure 3.8 represented by the OR
statements. The query also shows how the classification dates need to be explicitly considered to ensure that the lapse occurs after both the diagnoses.

### 3.4 Criteria Validation – The 20/20 Study

Being able to generate patient numbers for specific criteria on its own is of no importance unless the criteria can be validated to be useful by clinicians. In order to validate the criteria and also the audit report implementation, a validation study was initiated by the investigators. At the conclusion of the third meeting with the expert panel, eight specific quality improvement criteria were arrived upon as being the most important for the practice (Table 3.4) and formed the criteria that were to be used in the validation process; most of the details of this study have already been published in [17] and some relevant material is duplicated in this section.

<table>
<thead>
<tr>
<th>Table 3.4: The eight quality improvement criteria agreed with practice panel [17].</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of persistence of medication; and/or lapsed BP recording</td>
</tr>
<tr>
<td>1. A lapse in antihypertensive therapy &gt;30 days and the lapse extends into the Evaluation Period (EP)</td>
</tr>
<tr>
<td>2. A period of &gt;180 days with no BP measurements extending into the EP</td>
</tr>
<tr>
<td>3. A BP measurement of $\geq 160/100$ mmHg followed by a gap of &gt;120 days in BP measurements extending into the EP</td>
</tr>
<tr>
<td>Persistently high BP; lacking indicated therapy; and/or lab test contraindicating treatment</td>
</tr>
<tr>
<td>4. Three or more consistently high BP measurements ($\geq 160/100$ mmHg) over 120 days or more where either</td>
</tr>
<tr>
<td>i) the last of these high BPs was within the EP or</td>
</tr>
<tr>
<td>ii) with no subsequent “controlled” BP ($&lt; 160/100$ mmHg) measurements after the consistently high BPs</td>
</tr>
<tr>
<td>5. Classified with diabetes mellitus and not on ACEi/ARB at any time during EP</td>
</tr>
<tr>
<td>6. Classified with myocardial infarction and not on beta-blocker at any time during EP</td>
</tr>
<tr>
<td>7. Classified with renal impairment and on ACEi/ARB and with eGFR &lt; 60mL/min at any time during EP</td>
</tr>
<tr>
<td>8. On thiazide(s) and with serum uric acid $&gt; 0.42$mmol/l at any time during EP and not on Allopurinol or Colchicine</td>
</tr>
</tbody>
</table>

A more recent evaluation period not previously reviewed by the panel (during the iterative criteria development process) was used as the basis for a sample to validate the criteria. The criteria were assessed by database queries for the six-month evaluation period of 9 May to 8 November 2007 for all funded patients enrolled with the practice who had been classified with hypertension in the previous five years and had at least one antihypertensive prescription in the year prior to the beginning of evaluation period. A random sample of 40 cases total was drawn where:

- 20 cases were drawn from among those patients satisfying none of the eight criteria;
- 10 cases were drawn from those patients satisfying one or more of Criteria 1-3, but none of Criteria 4-8; and
- 10 cases were drawn from those satisfying one or more of Criteria 4-8 (irrespective of whether they also satisfied one or more of Criteria 1-3).

This process is illustrated in Figure 3.9.
There were 517 patients in the caseload for analysis. Of these cases, 209 (40.4%) met one or more of the eight criteria during the six-month evaluation period; 110 of the 209 (21.3% of total) met only Criteria 1-3, i.e., medication lapse or BP measurement lapse. Figure 3.9 shows frequencies with which the criteria groupings 1-3 and 4-8 were met.

The practice panel, independent of the external investigators and blind from the data query results, then assessed each of the 40 sample cases by review of records at the practice. For each sample case, the panel’s GPs then completed a three-question assessment instrument (Figure 3.10) which queried:

1. Freedom from significant contraindications or interactions;
2. Therapy optimised, or satisfactory process of seeking optimal treatment;
3. Adequacy of EMR data to support assessment of antihypertensive therapy.
After assessing all 40 cases, the criteria met by each case were revealed to the panel. The panel’s GPs then provided feedback and had the option to revise their assessment. Subsequently, there was a joint meeting of the panel GPs and external investigators to further consider the assessments and feedback on the sample.

Assessment of the 40 sampled cases by the practice panel yields six False Positives and six False Negatives (see Table 3.5) thus giving the observed validity for the sample as: sensitivity, 70%; specificity, 70%; positive predicative value (PPV), 70%; and negative predictive value (NPV), 70%; by chance the 95% Confidence Intervals (CIs) are the same for each statistic, (46-88%).
Table 3.5: Accuracy of automated queries as assessed against final review by the practice panel

<table>
<thead>
<tr>
<th>Classification by automated queries</th>
<th>Final classification by panel</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Met one or more criteria (+)</td>
<td>Quality suboptimal (+)</td>
<td>14*</td>
<td>6</td>
</tr>
<tr>
<td>Met none of the criteria (-)</td>
<td>Quality optimal (-)</td>
<td>6</td>
<td>14</td>
</tr>
</tbody>
</table>

* including three cases described as optimal on blind assessment but where practice GPs concurred to criteria on final review; all others initially negative on Question 2 (Figure 3.10)

Table 3.6 shows the frequencies of individual criteria in the caseload (i.e., 517 patients) and for the sample of 20 patients flagged as being on suboptimal therapy, and whether the panel concurred with the automated assessment. Note that three patients have moved from being False Positive to True Positive after the final assessment.

Table 3.6: Criteria and practice panel assessments for six-month evaluation period

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Number (as % of cases) in caseload</th>
<th>Number in sample</th>
<th>Number in sample where panel agreed with automated classification during blind assessment (after final assessment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>69 (13.3%)</td>
<td>7</td>
<td>4 (5)</td>
</tr>
<tr>
<td>2</td>
<td>79 (15.3%)</td>
<td>6</td>
<td>5 (5)*</td>
</tr>
<tr>
<td>3</td>
<td>21 (4.1%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>16 (3.1%)</td>
<td>1</td>
<td>1 (1)</td>
</tr>
<tr>
<td>5</td>
<td>39 (7.5%)</td>
<td>6</td>
<td>2 (2)</td>
</tr>
<tr>
<td>6</td>
<td>2 (1.0%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>25 (4.8%)</td>
<td>3</td>
<td>1 (3)</td>
</tr>
<tr>
<td>8</td>
<td>20 (3.9%)</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

* Criteria 1 and 2 co-occurred in two sample cases, one a True Positive and one a False Positive; Criterion 2 also co-occurred with Criterion 5 in a True Positive.

Table 3.7 and Table 3.8 report the specific panel feedback for False Positive and False Negative cases, respectively, with Table 3.7 including those three True Positive cases where the blind assessment was positive for Question 2 (“The therapy is optimised, or the process of seeking optimised treatment is satisfactory”).
Table 3.7: Cases where panel initially disagreed with classification by automated queries (False Positives); and final classification after review.

<table>
<thead>
<tr>
<th>Case</th>
<th>Criteria Satisfied</th>
<th>Panel Comments on Question 2</th>
<th>GP Comments upon viewing Criteria Satisfied</th>
<th>Final Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5: Classified with diabetes mellitus and not on ACEi/ARB at any time during EP</td>
<td>CVR 9%</td>
<td>On dietary management</td>
<td>False Positive</td>
</tr>
<tr>
<td>2</td>
<td>5: Classified with diabetes mellitus and not on ACEi/ARB at any time during EP</td>
<td>BP for diabetes CVR 10%; ACEi last prescribed Feb 2007, seems to have &quot;slipped off&quot;, list of regular medications</td>
<td>ACEi prescribed in Feb but not on list of medications (ACEi first prescribed in 2005)</td>
<td>False Positive*</td>
</tr>
<tr>
<td>3</td>
<td>1: A lapse in AHT of &gt;30 days during the EP or the lapse extends into the EP</td>
<td>CVR 3%</td>
<td>Patient has low CVR and may not need Rx</td>
<td>False Positive</td>
</tr>
<tr>
<td>4</td>
<td>1: A lapse in AHT of &gt;30 days during the EP or the lapse extends into the EP 2: A period of &gt;180 days with no BP measurements extending into the EP</td>
<td>CVR 6%</td>
<td>Needs BP treatments. CVR 6%</td>
<td>False Positive</td>
</tr>
<tr>
<td>5</td>
<td>1: A lapse in AHT of &gt;30 days during the EP or the lapse extends into the EP</td>
<td>CVR 13%</td>
<td>Not detected. CVR 13%</td>
<td>True Positive</td>
</tr>
<tr>
<td>6</td>
<td>5: Classified with diabetes mellitus and not on ACEi/ARB at any time during EP</td>
<td>CVR 15%; microalbuminuria; No uric acid recorded; BP not satisfactory</td>
<td>Given ACEi in 2003; only has IGT – misclassified</td>
<td>False Positive</td>
</tr>
<tr>
<td>7</td>
<td>7: Classified with renal impairment and on ACEi/ARB and with eGFR &lt; 60mL/min at any time during EP</td>
<td>CVR 8%; Chronic renal failure; No lipids or glucose measurements for four yrs</td>
<td>Agree</td>
<td>True Positive</td>
</tr>
<tr>
<td>8</td>
<td>5: Classified with diabetes mellitus and not on ACEi/ARB at any time during EP</td>
<td>CVR 19%; BP too high; Microalbuminuria</td>
<td>On accupril since Dec 2002</td>
<td>False Positive</td>
</tr>
<tr>
<td>9</td>
<td>7: Classified with renal impairment and on ACEi/ARB and with eGFR &lt; 60mL/min at any time during EP</td>
<td>CVR 20%; eGFR low; Cr high; BP too high</td>
<td>Agree</td>
<td>True Positive</td>
</tr>
</tbody>
</table>

EP = Evaluation Period; AHT – antihypertensive therapy; CVR – estimated 5 year cardiovascular risk as assessed by PREDICT [228], based on the Framingham risk equations with some adjustments specific to NZ; *Negative for Question 3 (refer Figure 3.10); IGT – Impaired glucose tolerance

Table 3.8: False Negatives

<table>
<thead>
<tr>
<th>Case</th>
<th>Panel Comments on Question 2</th>
<th>GP Comments upon viewing Criteria Satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Microalbuminuria reducing CVR 9%, but needs better BP control</td>
<td>BP too high; On thiazide + gout + uric acid 0.46</td>
</tr>
<tr>
<td>2</td>
<td>CVR 7%; Microalbuminuria improving. Poor diabetes control</td>
<td>Poor diabetes control*</td>
</tr>
<tr>
<td>3</td>
<td>CVR 15%; BP too high; Medication insufficient; Poor attendance</td>
<td>Poor attendance (late)</td>
</tr>
<tr>
<td>4</td>
<td>CVR 22%; BP Rx</td>
<td>Thiazide + history of gout + high uric acid; BP too high; CVR 22%</td>
</tr>
<tr>
<td>5</td>
<td>CVR 15%, poor attendance ACEi-cough; Statin – constipation</td>
<td>Poor attender; ACE-cough; statin not taken</td>
</tr>
<tr>
<td>6</td>
<td>BP could be lower; ACEi therapy not maximised</td>
<td>ACEi could increase</td>
</tr>
</tbody>
</table>

* On further consideration, practice ambivalent on False Negative / True Negative status of this case.

Clinical review on the sample of 40 cases appears to be moderately accurate indicators of cases relevant for follow-up. Review of the cases from Table 3.7 (False Positives) and Table 3.8 (False Negatives) reveals that the use of CVR within quality audit criteria may improve accuracy. The
collaborating practice was actually in the process of running PREDICT CVD/Diabetes [228] over all indicated patients, and thus had CVRs available to provide a convenient input to the criteria assessment process. Several observed False Positives (for cases 1, 3 and 4 in Table 3.7) had <10% CVR and several False Negatives had ≥15% CVR (for cases 3-5 in Table 3.8). In an ideal world the PREDICT CVD/Diabetes management recommendations could also be queried as part of the audit report queries, and also there would be access to more coded information on non-prescribing actions (e.g., the ‘on dietary management in Table 3.7, case 1). One could even envision running the PREDICT CVD/Diabetes algorithms in ‘batch’ mode to dynamically generate recommendations as part of the quality audit process, however this would only be effective if all required data were already in the EMR, which is often not the case (note: PREDICT CVD/Diabetes is usually run by manually entering any required information about the patient at run-time that is not already in the EMR). Other areas for improvement of the criteria performance related to more accurate interpretation of the prescribing record (noting cases 2 and 8 in Table 3.7, where re-prescription was done, but was missed by the database queries), less porous clinical criteria in specific areas (e.g., cases 1 and 4 in Table 3.8 around gout, uric acid and thiazides), and consideration of dose maximisation (case 6 in Table 3.8).

The diabetes/ACEi criterion (Criterion 5) seemed to be particularly vulnerable to the above problems and appears to be the most fruitful for refinement. It should be noted that the entire effort is underpinned by the high quality use of the PMS being exercised by the general practice, where the 40 cases reviewed indicated the records to be almost perfectly fit to purpose (with the Diabetes/IGT confusion in case 6 in Table 3.7 as the sole exception beyond the prescribing issues noted above).

Despite several limitations of this study, such as (1) using a single practice, where the coding practices, interests and biases of a few staff will have influenced the quality of the EMRs, (2) large confidence intervals for the aggregate of the criteria, and (3) small sample size to test individual criteria (indeed, not all criteria were observed in the sample), this study validated that EMR data can be used to provide a basis for moderately reliable automated identification of cases with suboptimal management of BP in the general practice setting. The criteria were shown to be suitable either for direct use in quality improvement efforts or for tracking of quality improvement outcomes over time. As such, the criteria were valid, but there was room for improvement, especially with regard to the implementation strategy.
3.5 Four Categories of Quality Audit Criteria

An outstanding challenge with the implementation discussed in Section 3.3 lies with the complexity of required query formulation and the technology of querying EMR data. There are some significant intermediate data structures that need to be built to answer questions as posed by the quality improvement criteria. These include:

- Identification of observations, both local observations such as BP and lab test results, which appear embedded in textual data fields, and which in the former case are sometimes, but not reliably, picked up by the PMS for coded representation in the EMR
- Grouping of medications into meaningful therapeutic groups (with consideration of issues such as combination drugs)
- Grouping of problem classification codes into meaningful groups
- The complexity of temporal queries considering the boundaries of an evaluation period and the order of events (e.g., continuity of ACEi after a diabetes diagnosis during a specific six-month period)
- Identification of the duration of a prescription (which is not always aligned to the period as stored in the practice EMR, depending on instructions given by the prescriber)
- Limitations around easily expanding the quality audit report to include other criteria as each query requires a criterion-specific query

As such, the development of queries by practice managers is impractical and one would need to compromise to ask these questions that are easily formulated rather than exploiting the true potential of the practice EMRs. Improvement of PMS query tools is required to suit the needs of quality improvement efforts. However, initiating any changes to a commercial system is not a trivial issue. This gives rise to the need for novel tools to be developed to support quality improvement efforts that could be used with EMR data, irrespective of any vendor specific system.

A key limitation of the MS Access–MS Word based implementation discussed in this chapter is the inflexibility around expanding the quality audit report to include other classes of criteria. For example, Criterion 1 involves identifying patients with hypertension who have a lapse in antihypertensive therapy for >30 days. The concept of hypertension is coded in the PMS using Read Clinical Codes, however, several codes can be associated with a hypertension classification. Similarly, the class of ‘antihypertensives’ is an abstract concept where multiple drugs (based on generic names and/or brand names) fall into the category of antihypertensive
medication. Therefore, it is necessary to maintain such knowledge via a more systematic approach than writing ad hoc SQL queries where all the required parameters (the Read Codes or the drug information in this case) are joined by OR operators. Irrespective of the specific SQL implementation, querying using such abstract concepts is challenging and calls for better knowledge management techniques – see next Chapter for an ontology based solution.

Other than abstracting specific concepts as discussed previously, there is also the opportunity to abstract the eight quality improvement criteria into more abstract classes of audit criteria - i.e., identifying the maximum reasonable level of abstraction required for prescribing for chronic disease based on EMR data, but excluding lifestyle modifications, surgery and the like. Based on this premise, four broad classes of audit criteria were identified: persistence to indicated medication (criteria 1, 5 and 6), timely measurement recording (criteria 2 and 3), time to achieve target (criterion 4) and measurement contraindicating therapy (criteria 7 and 8). Chapter 5 discusses these classes of audit criteria in greater detail.
Chapter 4

Preliminary Results – An Ontology Based Approach

The previously presented SQL based architecture was somewhat an ad hoc solution that could provide answers to a specific set of clinically important queries. However, this architecture was not generic enough to be easily extended to other domains and required much pre-processing with many intermediate steps. Therefore, a novel solution was required. This chapter discusses some of the drawbacks of the previous SQL based approach in some detail and then presents a novel, ontology based approach. The use of this new approach is demonstrated using electronic prescribing data from the same general practice mentioned in Chapter 3. A prescription timeline visualisation tool that has been developed is also discussed. Note that some material has been reused from one of my publications directly related to the work presented in this chapter [20].

4.1 Limitations of Using an Entirely SQL-based Approach (the Previous Implementation)

Most PMSs use relational database management systems to store patient EMR data. The SQL-relational combination inherently has a few limitations, especially in the chronic care domain. Three of the notable limitations are:

1. Lack of abstract, domain-level query support – if we want to write a query such as ‘number of patients in the practice with diabetes’ using a built-in reporting engine, we need to specify the explicit codes that constitute the concept ‘diabetes’ (as PMSs do not
provide specific relational tables constituting only the codes representing diabetes). Clinical coding systems (such as Read Codes, ICD-10 [229] and SNOMED-CT) tend to be vast, and generally provide a wide array of codes for closely related sets of concepts. For instance, in NZ the PHO Performance Management Programme [195] indicates 86 Read Clinical Codes to associate with a record of diabetes.

2. Lack of the notion of a hierarchy – the underlying relational data model has no inherent notion of a concept/query hierarchy. For example, simply because a particular drug is an ACEi/ARB drug, a typical SQL query engine cannot reason that this is also an antihypertensive drug. Ad-hoc workarounds such as using sub-queries are possible, but are error-prone and add complexity.

3. Nature of temporal SQL queries – SQL provides minimal temporal query support. This can be attributed to the lack of a formal temporal model in the underlying relational data model. Most PMSs provide an interface to query the underlying relational data using SQL, which has no temporal operators other than the simple before (via ‘<’), equal to and after (via ‘>’) on primitive date/time values. Therefore, even for a relatively simple query such as ‘patients classified with diabetes mellitus and not on ACEi or ARB any time during the evaluation period’ (i.e., with a lapse in ACEi/ARB supply) the corresponding SQL needs to satisfy the four cases illustrated in Figure 4.1.

![Figure 4.1: Temporal cases to be taken into account when formulating a query to determine ‘patients classified with diabetes not on ACEi/ARB any time during the evaluation period’](image)

Case 1 indicates where a patient was not on ACEi/ARB at the beginning of the evaluation period and then transitioned to having an ACEi/ARB prescription during the evaluation period. Case 2 is the opposite of Case 1 where the lapse occurs due to running out of supply.
during the evaluation period. However, we must also detect Case 3, where the lapse began before the evaluation period and extends into it. And these cases must be backstopped by Case 4 where there is no transition in ACEi/ARB prescribing – the diabetic patient simply received no ACEi/ARB medication at all. This fourth case is an extension from the three cases discussed previously in Figure 3.8 where the model was based on detecting ‘transitions’ in the presence vs. absence of a prescription providing medication supply.

Writing queries in SQL to support these types of requirements cannot be expected from a General Practice physician, and their construction is a challenge even for an experienced analyst. If such a query is in fact written, possibly with the assistance of a visual query builder tool, there is no guarantee of the validity of the results, as SQL will almost always return some result unless there is a syntax error.

4.2 Domain-Modelling and Methods

4.2.1 An Ontology-Based Approach

Further to addressing the three specific challenges discussed previously with regards to a traditional SQL-relational combination for querying, the following features are also desired from a possible (enhanced) solution:

- A sharable knowledge base – Clinical concepts such as hypertension or diabetes are domain-specific terms irrespective of the way they are coded within a PMS. In NZ for example, Read Clinical Codes Version 3 is widely used for coding patient classifications in General Practice, while in Australian General Practice WONCA’s ICPC is more popular. The US and the UK favour ICD-9 and ICD-10, respectively, in general, but coding schemes can vary within regions as well. Therefore, having a representation of domain specific concepts (rather than system-specific details) promotes shareability and reusability.

- An extensible knowledge base – This is required so that domain experts can easily refine the knowledge base to suit their needs. For example, it may be required to model diabetes not as a single concept, but as a hierarchy of clinical concepts to separate Type 1 diabetes, Type 2 diabetes and gestational diabetes mellitus. In the 20-20 study discussed in Section 3.4 and [17], high rates of kidney disease in the Pacific Island population gave particular priority to quality assessment at the intersection of hypertension and renal issues. A different practice context will have different drivers.
- Easy visualisation of knowledge base – Often with clinical audits, statements such as “patients with hypertension…” can be seen, however, it is not always obvious exactly what codes were used to define the condition (and hence what specific patient cohort is being audited); thus, having an easy visualisation/navigation capability is important.
- Standardised way of querying the knowledge base – Although SQL is an ANSI and an ISO standard, many database vendors support SQL with proprietary extensions to the standard language. Therefore SQL queries written targeting one database do not readily execute against a database from a different vendor even if everything else (tables names and so on) is the same.

An ontology driven approach was deemed to be a suitable solution that could satisfy all of the above requirements (the way it satisfied the last requirement will be discussed later). Ontologies are a crucial tool for formally specifying the vocabulary and relationship of concepts used on the Semantic Web [13], a technology/environment that provides structure for meaningful web content in order for software agents roaming from one webpage to another to readily carry out various sophisticated tasks for users. Therefore, having an ontological solution has the potential to facilitate easy integration with potential future developments in this domain. As discussed later, the ontology that has been developed is populated in the NZ context (with Read Clinical Codes only), but if a different coding scheme, such as SNOMED-CT, needs to be used, then the relevant segments of that scheme can be readily incorporated into the ontology. As such, the work presented in this chapter is essentially a Semantic Web Technology based solution with a domain specific ontology.

4.2.2 A Unified Patient Management Ontology

Different ontologies focus on different domains, sometimes even different views of the same domain. For example, several large ontologies have been developed in the medical/biomedical domains using Semantic Web technologies, most notably ProPreO [230] – a proteomics process ontology that models data provenance, process provenance, comparisons and analysis of proteomics datasets with over 3.1 million instances; and GALEN [231] – an ontology that contains building blocks for defining procedures, anatomy, surgical deeds, diseases and modifiers used in the definitions of surgical procedures. A more domain specific application that uses an ontology based approach is TrialWiz [232] which is an authoring environment that has been developed to manage the complexity of the clinical trial protocol-encoding process, improve efficiency in knowledge acquisition and perform related queries. TrialWiz implements an ontology-based specification of temporal information related to clinical trial protocols and
their application to the verification of protocol-specific temporal constraints among clinical trial activities. To process temporal information, TrialWiz uses technologies similar to those employed herein. A number of other active projects that use ontologies can be found at http://protegewiki.stanford.edu/index.php/Protege_Ontology_Library (accessed 10 October 2009).

My search for a suitable ontology compatible with the current domain of interest that could be reused for this work revealed that the ontology developed for hypertension management under the ATHENA decision support system [14, 103] was the closest match. This system is currently being used at the US Department of Veterans Affairs (http://www.chce.research.va.gov/athena.htm (accessed 10 October 2009). However, the ATHENA system is not currently publicly available and thus was logistically infeasible to reuse for the purposes of this thesis.

Despite some similarity of the clinical domain to that of some previous work, a fresh ontology was developed for the present demonstration taking the key concepts of relevant clinical guidelines (e.g., JNC7) in general, and focusing specifically on the eight criteria previously developed (as per Table 3.4). One key difference in modelling for quality assessment, as compared to interactive clinical decision support (as per ATHENA), is that the present concept hierarchies are organised to infer the relevant patient characteristics. For instance, ‘diabetic nephropathy’ is a diabetic complication (one affecting the kidneys). It is not the case that diabetic nephropathy is a kind of diabetes. However, a patient classified with diabetic nephropathy in fact has diabetes (with adequate certainty for quality audit purposes), and hence, for the specific purpose of the ontology discussed here, there is an effective is-a relationship (one might say a patient classified with diabetic nephropathy is-a patient classified with diabetes). Such application-specific approximations have been used to provide a more accurate audit even though the clinicians inputting the data were not working within the framework of an interactive decision support tool (i.e., they were free to code only as much information as they found relevant for their own notes).

With the assistance of the clinicians with whom we collaborate (primarily to identify the specific Read Codes they wanted to be included under the different classifications), I constructed a unified patient management ontology consisting of three main components: a disease management ontology, a patient data ontology and a best practice violation taxonomy (i.e., a taxonomy covering the required audit criteria) using the Web Ontology Language (OWL) [233].
Together with OWL, the Semantic Web Rule Language (SWRL) [234] was used as it allows users to write rules to reason about OWL individuals (i.e., the specific instances of a given type) and to infer new knowledge about these individuals. OWL has three increasingly-expressive sublanguages: OWL-Lite, OWL-DL and OWL-Full. SWRL supports only the first two sublanguages, and for the purposes of the ontology I developed, OWL-Lite was used due to its simplicity over the DL version, although either sublanguage could have been used. SWRL is a W3C standard and therefore satisfies the requirement that the solution should have a “Standardised way of querying the knowledge base” (per Section 4.2.1) – hence, these queries are more portable than vendor-specific SQL queries.

Protégé with the OWL plug-in (commonly referred to as Protégé-OWL) [235] was chosen as the preferred ontology development environment due to: (a) its active user base and level of available technical support; and (b) Protégé’s built-in support to create and execute SWRL rules. The main concepts of OWL are classes, properties and instances and it is herein assumed that the reader is familiar with these terms (see [233] for details on OWL constructs).

4.2.3 PMS Data Extraction

Using the same ethics protocol mentioned in Section 3.2 and the data extraction protocol in Section 3.3.1, a fresh dataset was extracted from the practice’s PMS, for the 18-month period ending 29th of January 2008 (with the exception of classifications which were extracted for five years back). The data extract involved 9710 patients, with 48651 prescriptions (13% antihypertensive medications), 11865 BP measurements and 31716 patient classification codes (encoding 1193 hypertension and 1041 diabetes mellitus classifications). The evaluation period was defined from 29th January 2007 – 29th January 2008 allowing a run-in period of 6 months so that patients who are already on therapy when they entered the evaluation period can be identified.

4.2.4 Patient Data Pre-processing and Dynamic Creation of OWL Properties

The patient data that was extracted from the practice’s PMS was pre-processed before populating the ontology created in Protégé-OWL (an overview of this process is shown in Figure 4.2). The built-in Java code generation feature Protégé-OWL has was used to provide a foundation for the data pre-processor.
The audit criteria shown in Table 3.4 require taking the different temporal relations shown in Figure 4.3 into consideration. OWL is based on open world reasoning, meaning that unless something is explicitly stated to be false, it cannot be assumed to be false. This means that unless it is explicitly specified what the next antihypertensive prescription is for a given antihypertensive prescription, there is no way in OWL to determine what this next antihypertensive prescription is, as OWL will assume that information that is not specified simply has not been added to the knowledge base yet. Therefore, instead of using an existing plug-in such as DataMaster [236] that supports direct data import/mapping into an ontology (thus creating the individuals in the ontology) from a database, I had to pre-process the data to explicitly create links to indicate what the next antihypertensive prescription is for a given antihypertensive prescription.
During the pre-processing stage, a number of T-Box (used to describe the relation between concepts/types) `next_prescription` OWL properties were created, so that a lapse in any antihypertensive drug class (such as ACEi/ARB, diuretic – refer to Figure 4.5 for the antihypertensive drug classes included in the ontology) can be identified just as easily as a lapse in antihypertensive. Using the relationships indicated by arrows in Figure 4.4, one can easily traverse the A-Box (which describes individuals rather than concepts) prescriptions for the drug class of interest. It should be noted that a drug class can have further subclasses (such as Loop Diuretics, Potassium Sparing Diuretics and Thiazide Diuretics under the Diuretics drug class), so the T-Box and A-Box `next_prescription` links are created dynamically based on the drug ontology for such subclasses as well. During investigation I found out that some prescription durations were longer than others issued on the same day (depicted by Diuretic Pr3 and ACEi Pr4 in Figure 4.4), and in such cases, the prescription with the longer duration was used as the link in the `next_AHT` chain.
Another issue that warranted careful consideration when creating the next_prescription links was to ensure that lapses at the beginning and end of the evaluation period could be properly identified (see Figure 4.3). A lapse running into the evaluation period can be identified when the last prescription prior to the beginning of the evaluation period is pointing to the first prescription during the evaluation period. However, to identify the lapses that occur at the start of the evaluation period (where there are no prescriptions prior to the evaluation period, but only a required classification) and the ongoing lapses it is needed to place an arbitrary prescription at the start and end of the evaluation period respectively (I decided to call these null_prescriptions). These null_prescriptions were created with a zero prescription duration. In terms of lapses, another possibility is when the patient has had no prescriptions at all during the evaluation period. Having no prescriptions at all in the current context (i.e., with an evaluation period of one year) may indicate the patient is no longer an active patient, but with a shorter evaluation period (say three months), patients with no prescriptions at all are important to identify for follow-up/recall. The two null_prescriptions created are used to identify these cases – if the null_prescription at the beginning of the evaluation period is pointing to the null_prescription at the end of the evaluation period, then the patient has had no prescriptions of required type (antihypertensives, ACEi and so on) for the entire duration of the evaluation period.

Similar to the next_prescription links and the null_prescriptions that were created for prescriptions, next_measurement links and null_measurements were created for each measurement/lab result type as well (i.e., for BP, creatinine and so on – refer to Figure 4.7 for measurement/lab result types included in the ontology).
4.3 Identifying Patients with Hypertension on Suboptimal Therapy using the Ontology Based Approach

This section described how the unified patient management ontology (consisting of a disease management ontology, a patient data ontology and a best practice violation taxonomy) has been used to identify hypertensive patients whose therapy can be improved and describes the functionality of the main components of the proposed approach.

4.3.1 Disease Management Ontology

Figure 4.5 shows the knowledge base classes constituting the disease management ontology, created using Protégé-OWL (version 3.4 – Build 130). The previously determined criteria (Table 3.4) provided guidance on the nature of clinical concepts that needed to be included here. In this case, the ontology focuses primarily on antihypertensive therapy, but can be easily extended to other domains. In the domain of hypertension, one also has to consider comorbidities (concurrent conditions) that are relevant to antihypertensive therapy management; for example the presence of asthma contraindicates the use of beta-blockers which is otherwise a first-line BP lowering medication; therefore common comorbidities related to antihypertensive therapy (asthma, diabetes mellitus, myocardial infarction and renal impairment) are also included. All instances of this ontology were created using xsd:string datatype [237].
The grouping and naming of the ontology concepts is mostly self explanatory but a few details need to be explained. The class *ClassificationCodes* has been designed to accommodate the commonly used coding schemes Read Clinical Codes, SNOMED, ICD-9 and ICD-10, although only the *ReadClinicalCodes* class has been populated since this is the classification coding scheme used in NZ. These classes represent a pool of classification coding terms and the idea is to define the subclasses of *ProblemClassifications* using the classification codes from this pool. For example, the concept *RenalImpairement* can be restricted (via OWL properties) to consist of the selected members from the *ReadClinicalCodes* class (as well from any of its sibling classes). This approach ensures consistency of the classification terms ('hypertension' for example) irrespective of the coding scheme used by a region and/or vendor specific system.

With the assistance of the clinicians we collaborate with, 63 Read Codes were identified constituting the definitions of asthma, diabetes mellitus, hypertension, myocardial infarction and renal impairment (Figure 4.6). In a very few cases, the same Read Code has been used to
identify multiple classifications – Read Code for diabetic nephropathy is C104.11, which is a diabetes mellitus classification as well as a renal impairment classification.

The class Drugs is used to hold information about all the raw drugs based on their generic names and has been populated with 58 antihypertensive drugs as relevant to the current context. It is from this pool of drugs (more formally, instances of the class Drugs) that the other drug classes (ACEis, Diuretics and so on) are populated. The process is analogous to defining the various Classifications by selecting the required Read Codes from the ReadClinicalCodes class. The previously mentioned null_prescription is an artificially introduced one, so a null_drug instance (created as an instance of SpecialDrugs class in Figure 4.6) is also introduced – each patient prescription has an associated drug selected from the Drugs class and this null_drug was used for the null_prescriptions.

Members of both classes, ProblemClassifications and DrugClasses may belong to multiple parents hence having a multiple inheritance structure in the ontology. From a strict philosophical and engineering perspective this may not be considered best practice ontological modelling, however, returning to the example of diabetic nephropathy (with Read Code C104.11), it is-a
diabetes mellitus Read Code and also is-a renal impairment Read Code. Combination drugs provide another case for multiple inheritance - for example, the drug ‘quinapril hydrochlorothiazide’ (usually prescribed as single combination tablet under the popular brand name Accuretic) is-a ACEi and also is-a thiazide diuretic. Using this type of a multiple inheritance structure is consistent with systematised medical terminology such as SNOMED-CT which classifies diabetic nephropathy (SNOMED concept id 127013003) as a renal disorders in systemic disease as well as a diabetic complication; and quinapril hydrochlorothiazide (SNOMED concept id 421852001) as an angiotensin-converting enzyme inhibitor agent and a hydrochlorothiazide.

4.3.2 Patient Data Ontology

The patient data ontology represents the concepts related to a patient. It holds information about the classifications a patient has, prescriptions, measurement results (only BPs in this case) and various laboratory test results (eGFR, uric acid and so on). xsd:float datatype [237] was used to represent all measurement and test result values while ValidPeriod and ValidInstant built-ins from the SWRL temporal ontology [238] were used to represent prescription durations and measurement/test dates, respectively. These SWRL built-ins were used for temporal representations over other W3C temporal concept ontologies such as OWL-Time [239] due to their out-of-the-box support from Protégé and use in other similar applications [240].

The main interest of this work was in patients who were enrolled and funded (i.e., the practice is responsible for managing these patients – see Section 2.4) and had at least one hypertension diagnosis. Based on these filter criteria, out of the extracted 9710 patients, only 538 were populated into the ontology during pre-processing (Figure 4.7). The final populated ontology consists of over 15,000 individuals, 60 classes, 30 object properties and 15 data type properties, and the Protégé screenshots in Figure 4.6 and Figure 4.7 provide illustration of some of the classes and individuals. The Protégé project with the OWL ontology file is available online via http://www.cs.auckland.ac.nz/~thusitha/aiim09/.
It is worth noting in Figure 4.7 that laboratory results and BP measurements have two *null_measurements* (for this particular patient) while prescriptions have two *null_prescriptions*. As discussed earlier, these are used to mark the temporal boundaries of the prescriptions, BPs and laboratory test results a patient has had for the evaluation period of interest.

### 4.3.3 Best Practice Violation Taxonomy

Shown in Figure 4.8 is the explicit, best-practice violation taxonomy that has been developed which is representative of the eight audit criteria developed previously, as well as the four broad categories of quality audit criteria discussed in Section 3.5.
All subclasses of \textit{Criteria} starting with $C_\ast$ in Figure 4.8 are various \textit{generic} clinical audit concepts. For example, out of the eight audit criteria, the fifth was ‘Classified with diabetes mellitus and not on ACEi/ARB at any time during the evaluation period’ and this criterion is represented in Figure 4.8 by $C_5\_PatientsWithDMNotOnACEiARB$. This is included under the generic criterion $C\_LapseOccurringAfterIndicatedClassification$, which itself ‘is-a’ $C\_LapseInTherapy$ criterion. Having this type of a taxonomy hierarchy provides flexibility to easily add any new audit criteria as required. The class $AnalysisScope$ is used simply to hold properties that are applicable across all criteria, such as ‘all patients should have a hypertension classification’. The class $CriteriaRelatedSubQueries$ contains intermediate query concepts, such as the query to determine whether a patient is hypertensive or not, as required by various other queries. Having these intermediate query concepts promotes reusability and modularity by removing duplication of queries. This ontology is not instantiated by manually creating individuals, but populated automatically with the corresponding individuals as the result of executing a SWRL rule, hence a specific datatype cannot be specified.
4.3.4 Querying the Ontology using SQWRL

As shown in Figure 4.9 and Figure 4.10, the Protégé-OWL editor has support to edit and execute SWRL rules via its SWRL Rules tab. The queries involve determining the number of patients who satisfied a certain criterion, and it was preferred to display the query results ordered by certain fields such as the date of prescription. Therefore, the Semantic Query-enhanced Web Rule Language (SQWRL) was used which is a SWRL-based query language that can be used to query OWL ontologies and also provides SQL-like operations to format knowledge retrieved from an OWL ontology [241]. SQWRL queries can be written and executed on the same SWRL Rules tab, and have additional self-explanatory constructs (i.e., compared to SWRL) such as orderBy, selectDistinct and countDistinct. It should be noted that SQWRL is designed to be a query language only, whereas SWRL can be used to create new knowledge.

Protégé-OWL has a powerful SWRL temporal built-in library capable of representing temporal durations (used to represent prescription durations in our case) as well as perform various other temporal operations based on the standard Allen temporal operators [238]. SQWRL implements most of these operators backed by the Jess rule engine [242] and provides an ideal platform to write the required temporal queries for the purposes of this work. The corresponding SQWRL query to identify hypertensive patients with a lapse in their antihypertensive therapy for over 30 days with lapse extending into or occurring during the evaluation period (i.e., the first audit criterion) is shown in Figure 4.9 (for syntax and other SWRL and SQWRL related details refer to [241]). This query primarily queries the patient data ontology while using constructs from the SWRL Temporal Built-in Library [238]. Note that here the query is being executed to view the results (using sqwrl:selectDistinct construct), while if needed, the query can be easily converted to a SWRL rule and the results inserted into the corresponding class in the best practice violation taxonomy (i.e., C_LapseOccurringAfterIndicatedClassification in this case).
Figure 4.9: SQWRL query to identify hypertensive patients with a lapse in their antihypertensive therapy for over 30 days with lapse extending into or occurring during the evaluation period

Figure 4.9 shows how a number of individually simple SQWRL assertions can be placed into a logical conjunction to formulate a query expressing multiple temporal constraints. Using SQWRL in this manner provides the capacity to create queries with higher level domain concepts, making the construction of queries fairly straightforward as compared to corresponding SQL. Furthermore, the resulting syntax looks much simpler and eliminates the need for the query author to be concerned with implementation details such as specifying the JOINs of multiple database tables. The various other queries written to identify the patient cohorts belonging to the remaining audit criteria are shown in the top half of Figure 4.10.
Figure 4.10: The rule/query authoring and executing environment. The top half shows the various SWRL rules and SQWRL queries that have been written while the bottom half shows the results after executing the highlighted query.

The bottom half of Figure 4.10 shows the results of executing the query shown in Figure 4.9 (together with the helper sub-query PatientsWithHT_qry). The query results indicate the patient identifier and details related to the lapse in therapy that occurred. For instance, the highlighted portion of the query results shows that the patient with identifier M012237 was classified with essential_hypertension (with Read Code G20.00) on 12th July 2005, and had three antihypertensive lapses of over 30 days extending into or occurring during the evaluation period: i) 10th February 2007 – 28th March 2007 (46 days), ii) 19th August 2007 – 21st September (33 days) and iii) 20th December 2007 – 29th January 2008 (40 days and is an ongoing lapse). For technical quality assurance, the cases identified by the SQWRL approach were compared to those identified with the previous SQL based approach for the established criteria and ensured that they were identical.
4.3.5 An Excel-based Prescription Timeline Visualisation Tool

The concept of medication ‘lapse’ has been discussed so far using a lapse duration, measured in days. This is derived by analysing prescribing data and the raw prescription data contains only the prescription date and duration (among other fields). This makes determining lapse durations for a given set of prescriptions a difficult task, especially by visual inspection of the prescriptions in the database (Figure 3.5 shows some prescription information). Some patients are on multiple agent therapy which further complicates this analysis. It is possible to verify that the current ontology based implementation of identifying lapses is correct perhaps using another independent implementation and then comparing the results, however, validating that the lapses (durations, medications, from/to dates and number of lapses) are what they really should be is difficult. My first attempt towards ensuring that the lapses were correct was to select a random patient (patient M012237 from Figure 4.10 for example) and then write down on paper the different prescriptions the patient had had along a timeline and then manually determine the lapses and compare those to what had been identified by SQWRL. Having the prescriptions placed on such a timeline also helped in getting a ‘feel’ for the nature of lapses a given patient is having; for instance, the visual representation can indicate whether the patient is a frequent late comer for the next prescription.

The above mentioned manual visualisation of prescriptions was a tedious task consuming much time. Driven by the need for a more automated visualisation scheme, I investigated different charting utilities, but it appeared that my charting requirements were different from what was freely available. The placement of prescriptions along a time axis has some commonality with standard Gantt Charts, but still different in terms of the layout. Therefore, using any Gantt Chart plotting techniques was not an option. After evaluating several options, Excel was chosen as the development tool and used with its VBA modules to read a specified text file with prescription information and then plot the required chart. This prescription file was created during the data mapping stage depicted in Figure 4.2 where local database information is transformed into Protégé. Figure 4.11 shows the resulting prescription timeline visualisation tool for patient M012237 from Figure 4.10.
Having this type of a tool makes it easier to validate that indeed there are only three >30 day lapses for this patient as well as the prescription/lapse locations, as indicated by Figure 4.10.

4.4 Other Ontology Related Work and Limitations of a Pure Ontology Based Approach

This chapter presented a novel approach that can be used to enhance the querying capabilities of general practice medicine in the context of chronic disease management, focussing mainly on hypertension and its comorbidities. The nature of chronic disease requires reasoning on time intervals, which are highly implicit in PMS databases that store time-stamped events. Thus, capturing the clinical requirements necessary to characterise good chronic disease management leads to highly-complex and potentially error-prone queries. The approach presented herein can be used to cater for such domain-specific querying requirements of quality improvement for general practice activities, such as treatment of hypertension. Furthermore, the use of an ontology based approach makes it easier to create, visualise and navigate the knowledge base (the specific Read Codes that constitute diabetes mellitus, for example) making it a more clinician-friendly approach.

A key strength of the presented approach is its modularity and extensibility – the three components of the unified ontology can be easily extended to include other clinical concepts, as well as add new query capabilities. Including provisions for such future expansions is worth mentioning, for example, including SNOMED, ICD-9 and ICD-10 placeholders for problem classification whereas only Read Clinical Codes have been populated in the NZ context.
Despite the previously discussed issues with relational databases (see Section 4.1), it should be possible to use existing relational database technologies (with multiple UNION type SQL queries and/or sequential data processing), a rule base and intelligent scripting to create a framework with similar functional capability to the one presented herein. However, Protégé already provides a clean user interface capable of modelling domain level concepts with easy visualisation and hierarchy navigation capabilities, providing an excellent opportunity to use existing tools instead of developing a custom-built user interface. Furthermore, the temporal operator library that has been used in this work is available as part of the out-of-the-box Protégé installation, which means there is no need to process the relational data using purpose-built temporal operators that would need to be implemented. As such, Semantic Web technologies already have the required tools to model the requirements that have been set forth, and hence this was the preferred approach.

During the preliminary stages of this research, I considered creating a layer on top of the relational database model to support temporal querying with better expressivity by using a temporal database mediator such as Chronus-II (see Section 2.3.6.1 and [152]), developed at Stanford Medical Informatics. However, after some discussion with my supervisor, we decided to proceed with a Semantic Web technology based approach for several reasons: it is one of the newer, evolving technologies with an active user community worldwide; there is an expanding array of support tools freely available; and because the Chronus-II project is no longer active project at Stanford (otherwise it could possibly have provided a base for this architecture).

The present proposition focusses chiefly in the emphasis on analysing patient management over a time period and with application to quality improvement over a cohort of patients rather than as interactive decision support at the point of care (which has been the main focus of other automated-guideline based CDSSs such as PRODIGY and ATHENA; see Section 2.2 for details). In fact, this approach is more like IDAN/KNAVE-II temporal abstraction/visualisation framework [143]. However, the current ontology based framework is developed using Semantic Web based technologies and driven by a tight focus on domain specific clinical requirements of prescribing for chronic disease management rather than generic data abstraction and exploration.

SWRL has been used to write the inference rules required and SQWRL for querying the knowledge base. SWRL is a rule based language that extends the expressivity of OWL at the
expense of theoretical decidability of query answering operations [243], although this is never really an issue in most practical applications. OWL also supports efficient description logic (DL) based reasoning, however querying the reasoner inferred knowledge using SQWRL (or any other OWL-aware mechanism) is currently not possible (as of Protégé-OWL build 130). There are plans to integrate the Pellet DL reasoner with SWRL support with Protégé [244] (as of build 130, there is no direct integration with a reasoner). As such, an alternative future implementation of the current querying mechanism may consider using a DL reasoner based technique instead of the rule-based approach presented herein; for example, it is possible to write DL statements to define a patient with hypertension as a patient who has a classification where the code is a hypertension classification code. A DL reasoner can then infer all patients with hypertension. SWRL rules may also be used in conjunction with these DL statements when additional expressivity is needed, thereby resulting in somewhat a hybrid solution. Another important future consideration may need to look at developing a testing framework that can be used to verify the validity of the SWRL rules and the SQWRL queries that have been written – some verification has been done in terms of comparing the results with those from the previous SQL based approach, and observing improved BP in adherent patients as discussed in [21] provides some validation, however, a more systematic approach would be more desirable.

My experience on SQWRL based querying is that in general, it is much slower than an equivalent SQL based implementation (where the query engine is usually highly optimised for data retrieval), with some of the SQWRL queries taking several minutes to generate results as the dataset gets larger (for example, >100,000 OWL individuals in total consisting of patient details for prescriptions, classification codes and BP measurements). However, in the current context, where we are envisioning a periodic quality assessment analysis (as compared to an interactive decision support application at the point of care), this has not been a significant drawback. In fact, a more considerable limitation can be observed from a logistic/maintenance point of view – as Figure 4.2 shows, currently a number of pre-processing steps are required to prepare the data to be in a queryable format. Also, it is not ideal to import all data into Protégé because of the number of individuals that would need to be created (i.e., patient, prescriptions, classifications, various measurements and so on). Theoretically this may not be an issue, but due to limited computer power (primary limitation being random access memory), having a large ontology is problematic. In an attempt to overcome these issues, it was desirable to maintain all patient data in a relational database (that is already optimised for handling large volumes of data, as well as querying) while using an ontology based approach to maintain the knowledge bases to make use of the advantages discussed in Section 4.2.1. Discussing an
enhanced architecture with this separation of data and knowledge is the focus of the next chapter.
Chapter 5

ChronoMedIt Framework Architecture

The ontology based approach discussed in the previous Chapter excelled in terms of providing a viable solution to maintain the knowledge base. However, as discussed towards the end of the Chapter, there were performance issues with respect to querying speeds because of the large number of patient records in the ontology. On the other hand, the SQL based architecture discussed in Chapter 3 was efficient in querying, but knowledge maintenance was difficult. Naturally, this led to developing a framework taking a hybrid approach composed of using SQL based techniques for querying purposes while maintaining the required knowledge bases using an ontology based approach. I decided to call this resulting framework ChronoMedIt (indicating Chronological Medical audit) as the framework takes various temporal considerations into account when formulating and executing audit criteria as relevant to chronic disease management. ChronoMedIt is an important result of this thesis, and this chapter presents details pertaining to its architecture and implementation. Note that some material has been reused from a related publication [23].

5.1 The Need for a Novel Criteria Model

An inherent limitation of the past approaches is that the implementation was based on developing ad-hoc type queries (irrespective of the querying technology used) to provide answers to clinically important queries, but without the criteria being grounded in an underlying criteria model. The lack of a criteria model has the following disadvantages:
1. Interpretation of criteria formulated using natural language may be ambiguous while a methodologically developed, well-defined criteria model promotes logically crisp interpretation.

2. Extending the existing set of criteria (per Table 3.4) becomes a difficult task and may require writing an entirely new query, when it may have been possible to use a simple extension to an existing criterion.

3. Verification of criteria becomes a challenge as each individual criterion needs to be verified accordingly. If common criteria are present, verifying the common criteria once eliminates the requirement to verify the common aspects of each specific criterion (which requires a greater effort).

Therefore, it was required to develop a criteria model to overcome these issues. The eight criteria that have been discussed thus far were used as exemplars to guide development of this new criteria model. The model was to have the flexibility to formulate a range of audit criteria that could be easily extended as required. The framework implementation was to closely follow the criteria model so that any additional criteria to be modelled can be easily implemented. As discussed in some detail in Section 4.1 and Section 4.2.1, the other main requirements included criteria formulation using domain-level therapeutic and classification knowledge, using a shareable knowledge base (for drugs and classification details) and easy visualisation of the knowledge bases.

5.2 Framework Verification

The development of a medical audit framework is not sufficient in itself; as with any robust software system, it is important to develop a mechanism to verify that framework output results are accurate. Therefore, it was required to develop a scheme whereby criteria output results could be verified. The approach taken herein to verify the framework implementation was threefold:

1. Develop two independent implementations (using different algorithms) to determine patients satisfying a given criterion so that the results from the two implementations can be compared.

2. Develop test cases to perform boundary value analysis, equivalence class testing and all-pairs testing for various test cases that cover the breadth of possibilities resulting from the Criteria Model that will be developed (as per Section 5.1). The next chapter discusses these testing techniques in some detail, but refer to [245, 246] for a detailed discussion.
3. Perform random testing to ensure software reliability – Adequately testing any software system of reasonable complexity for all possible inputs requires a great effort and the number of test cases required can easily become unmanageable. Random testing is a form of software testing that refers to selecting test inputs randomly from the input domain of a system [246]. Performing random testing with a large number of test cases has the potential to cover all possible combinations of the input domain (with some probability for each combination to occur), and has been suggested as a possible solution to ensure software reliability [246, 247]. A concept closely associated with generating random test cases is the operational profile of the input space [247, 248] which is simply a set of disjoint (only one can occur at a time) alternatives with the probability that each will occur [248]. The operational profile should take into account usage patterns the software will encounter in its intended environment. The argument here is that by testing according to usage, the failures found by imperfect methods are more likely to be the important ones, that is, the ones most users would first encounter [247]. It has been suggested that when defining this profile, a practically complete list of input variables from the input space should be chosen [248], and therefore a suitable domain specific operation profile was to be determined when generating the test cases.

Since framework testing constitutes an integral part of my research, Chapter 6 is devoted to discussing testing concepts as well as the test cases that have been developed.

5.3 Introducing Medication Adherence as an Audit Criterion

Research has shown impressive efficacy rates of BP lowering medications for reduction in cardiovascular and renal events. These antihypertensive medications come in a variety of classes and subclasses, with varying indications and also used in combination, depending on the characteristics of a patient, such as the presence of diabetes or heart failure. Although these drugs are effective when taken as directed, low adherence (also sometimes referred to as ‘non-compliance’) to prescribed regimens threatens their effectiveness in real-world use. Long-term adherence with medications for chronic diseases is low in general, particularly among lower socio-economic groups [249]. Based on the international literature, it is thought that poor adherence to antihypertensive medication contributes to inadequate BP control in more than two-thirds of hypertensive patients [250]. A Swedish study found satisfactory refill adherence for thiazide diuretics at 55%, ACEi at 59% and selective beta-blocking agents at 66% [251]. The level of adherence has been positively correlated with good BP control and reduction in the complications of hypertension [252]. Low adherence has been cited in the international
literature as the primary cause of unsatisfactory control of BP [6], however, studies that focus on identifying specific cohorts of patients who are non-adherent are limited. In light of the large magnitude of long-term medication adherence issues, and the ongoing need to formulate successful adherence improvement strategies, including a measure of medication adherence as an audit criterion was considered important and the framework was to provide support for relevant adherence queries as well. In fact, the first audit criterion out of the eight that were identified is directly related to medication adherence and was based on clinician intuition as an important issue at the practice that had not been addressed previously.

Andrade et al. [253] in their systematic literature review that investigated different adherence measures concluded that identification, aptness and selection of measures for adherence should be determined by the objectives of the study, and limitations and benefits of the measures should be considered. Adherence refers to the extent to which a patient’s behaviour to take the prescribed medications aligns with the instructions and recommendations from the prescriber [254]. A widely-used measure of adherence is defined in terms of a proportion-of-days-covered model, which calculates the proportion of days within a fixed interval that the patient has an available supply of medication [255], and reported using the medication possession ratio (MPR) [253].

In this thesis, I have employed persistence based adherence measures of proportion-of-days-covered based on the prescribing of the patient’s general practice. This is a type of an MPR measure, although alternatively the MPR may be computed from dispensing or by direct observation of pill counts at the patient’s home. MPR is often expressed as a percentage and an MPR of less than 80% has been indicated as a threshold upon which clinical effectiveness of a therapy is significantly compromised [253]. A patient is usually referred to as ‘non-adherent’ if their MPR is less than 80%. It should be noted that the first criterion per Table 3.4 (i.e., lapse in therapy for >30 days) is also an adherence measure, but focusses more on significant individual lapses whereas poor MPR (i.e., MPR <80%) is an indication of overall medication adherence for a given period. Some authors [253] distinguish between adherence and persistence where persistence generally indicates whether a patient continues to be on therapy from initiation to discontinuation of therapy. The first audit criterion relates to continuation of therapy and from a terminology perspective may be considered a persistence measure. However, for the purposes of this thesis, I have not made a strict distinction between these two terms as the goal is to identify patients who have had poor possession of medication, irrespective of whether they occurred as significant disjoint lapses or multiple short lapses (resulting in low MPR).
After reviewing 77 studies, Andrade et al. [253] reported that the two most frequently used definitions for MPR are:

\[
\text{MPR} = \frac{\text{# of days supply obtained during observation period}}{\text{# of days in observation period}} \times 100
\]

(1)

and

\[
\text{MPR} = \frac{\text{# of days supply obtained (excluding last refill)}}{\text{# of days between first and last dispense dates}} \times 100
\]

(2)

Based on my experience in analysing prescribing data, to determine the MPR for patients after a required classification (such as hypertension), there are several temporal considerations (Figure 5.1) that need to be taken into account for an accurate measure of MPR.

**Figure 5.1: Temporal relations to consider when calculating MPR. Prx denotes a prescription where x is a sequence number. ● represents a patient classification and TotGap indicates the total lapse in medication that needs to be considered when determining medication possession during the period of interest. Scenario (i) is a typical case where the medication lapse occurs during the evaluation period (EP); (ii) is a case with a lapse running into the EP; (iii) is an on-going lapse at the end of the EP; (iv) is a case where only part of the total lapse is needed; and (v) and (vi) represent cases where there is no medication at all during the EP, but in (vi), only part of the EP needs to be considered.**

It needs to be acknowledged that the definitions in equations (1) and (2) are commonly used for MPR and may be adequate for the majority of cases, however, in this research I attempt to offer...
a more refined definition that considers a patient’s medication availability at the edges of the evaluation period, as well as the time of classification. For example, consider scenario (i) in Figure 5.1. If definition (2) is used to calculate MPR for this prescribing scenario, only the second prescription (denoted by Pr2) will be included, which is a rather incomplete picture of medication possession for this patient over the evaluation period. In terms of the definition in (1) for scenario (i), failure to account for Pr1 (because it was not ‘obtained during the observation period’) tends to underestimate medication possession; conversely, the medication possession from Pr3 will be overestimated. Moreover, if a denominator of the entire duration of the evaluation period is used for a scenario such as (iv) where Pr2 is the patient’s first prescription of that drug after being classified, the resulting MPR would be misleadingly low. If a period of considerable duration (such as a year) is considered, there are likely to be many patients who are newly diagnosed during the evaluation period, and, as such, if low MPRs are used as the basis to identify non-adherent patients, cases with a prescribing pattern similar to scenario (iv) will contribute towards false-positives resulting in low specificity. Similar issues arise when definition (1) and/or (2) is applied to other scenarios.

Therefore, for the work presented herein, MPR is calculated by including the boundary prescriptions such that if scenario (i) in Figure 5.1 is considered, only those parts of Pr1 and Pr3 coverages that fall within the evaluation period are included in the numerator of the MPR calculation. Furthermore, a run-in period prior to the evaluation period is used (as discussed previously) so that prescriptions such as Pr1 (prescribed prior to the beginning of the evaluation period that run-into the evaluation period) can be correctly accounted for.

The following definitions have been used for MPR calculations:

If patient was classified before the beginning of the evaluation period (EP):

\[
MPR = \frac{\# \text{ of days in EP} - \text{Total gap duration}}{\# \text{ of days in EP}} \times 100
\]  

(3)

Else:

\[
MPR = \frac{(\# \text{ of days between classification date and end of EP}) - \text{Total gap duration}}{\# \text{ of days between classification date and end of EP}} \times 100
\]

(4)

where Total gap duration refers to the sum of all medication lapses as determined after various temporal considerations shown in Figure 5.1.
It should be noted that several possession based adherence measures have been identified in the literature. A recent study by Karve et al. [256] compared 11 different adherence measures based on a systematic literature review on long-term medication adherence measures by Hess et al. [257]. The authors indicate that two measures —proportion of days covered (PDC) uncapped and PDC capped (at a maximum of 100%) should be considered first when selecting among different adherence measures. The definitions for these two measures (adapted from Karve et al. [256] and Hess et al. [257]) are shown in Table 5.1.

| Table 5.1: Definitions of PDC uncapped and PDC capped |
|---------------------------------|------------------|
| Adherence Measure          | Formula                                     |
| PDC uncapped               | \( \frac{\text{Number of days supply during EP}}{\text{Number of days in the EP}} \times 100\% \) |
| PDC capped                 | \( \frac{\text{Number of days supply during EP}}{\text{Number of days in the EP}} \times 100\% \) at capped 100% |

The definition of MPR is somewhat similar to that of PDC capped since the maximum value of MPR is 100% which is the case when the patient is fully adherent. MPR does not account for medication oversupply explicitly so there is no room for gradual accumulation of medication (i.e., stockpiling) resulting in a reasonable period of oversupply which can shadow non-adherence. Further, I have included the special case (v) in Figure 5.1 so that patients who have not been on any medication at all during the evaluation period can also be identified.

The two adherence measures in Table 5.1 use the entire duration of the evaluation period as the denominator and studies that use these definitions for adherence calculations usually include only patients who were classified prior to the evaluation period. “Most patients with hypertension will require two or more antihypertensive medications to achieve goal blood pressure” [6]; however most studies only consider patients on monotherapy in an attempt to simplify adherence calculations [256, 258-260] and this is likely to exclude patients with greater disease severity [256].

Figure 5.2 shows an example scenario where values produced by each of the measures can be directly compared. For comparison purposes, assume a hypothetical patient who was classified with hypertension prior to the beginning of the evaluation period. Also assume that the patient was initially on ACEi monotherapy (denoted by Pr1, Pr2, Pr3 and Pr5 prescriptions) and then a diuretic was added during the evaluation period (denoted by Pr4 and Pr6) to intensify therapy.
Figure 5.2: Prescribing patterns for a hypothetical patient. The gray rectangular boxes denote prescriptions (Pr) and the numerical values within a box indicate the prescription duration. Relevant temporal durations required for adherence calculations are also shown. All durations are shown in days.

Table 5.2 shows the overall antihypertensive medication adherence rates based on the different measures.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Adherence Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDC uncapped</td>
<td>90 x 5 / 365 = 123%</td>
</tr>
<tr>
<td>PDC capped</td>
<td>90 x 5 / 365 [capped at 100% if PDC&gt;100%] = 100%</td>
</tr>
<tr>
<td>MPR</td>
<td>[365 - (90 + 35)] / 365 = 82%</td>
</tr>
</tbody>
</table>

It should be clear that the patient in Figure 5.2 is not fully adherent to medication as suggested by PDC capped and PDC uncapped. The scenario in Figure 5.2 is a fairly realistic ‘real-world’ example and the purpose of this was to make the reader aware of some of the important temporal relationships that need to be considered. If the end date of Pr1 in Figure 5.2 was before the beginning of the evaluation period for example, there would be a treatment gap at the beginning of the evaluation period as well, but PDC capped and PDC uncapped would still result in the same values whereas the MPR calculation used in this thesis would result in a more realistic value of [365 – (90 + 35)] / 365 = 66%.
5.4 The Criteria Model

Using the eight criteria developed previously as guidance on the requirements to engineer, a generic criteria model was developed using UML. Four broad, distinct criteria classes (also referred to as elements) were previously identified as sufficient to cover the requirements (see Section 3.5) and the resulting UML model is shown in Figure 5.3. Note that the not on Allopurinol or Colchicine clause in the last criterion per Table 3.4 has been ignored in the criteria engineering effort. This is because discussions with our collaborating clinicians have indicated that patients having high serum uric acid levels while on thiazides need to be identified irrespective of whether they are on Allopurinol or Colchicine.

![Figure 5.3: Elements of ChronoMedIt’s Criteria Model (using UML notation). Unless the multiplicity of a given attribute is specified ([0..1] for example) a multiplicity of 1 is assumed (i.e., a compulsory attribute).](image)

The MedicationLapse element in Figure 5.3 is used to formulate criteria that are primarily related to persistence to indicated medication (criteria 1, 5 and 6). Although Table 3.4 does not have a direct MPR related criterion, I have included the possibility to create adherence related measures in the Criteria Model – for example, if it is needed to identify patients with MPR <80% (a commonly used MPR threshold to identify patients with poor adherence [253]), MedicationLapse can be specified using maxMPRThreshold = 80 and compareMaxMPRUsingOp = LESS_THAN. If minLapseDuration attribute was also set to 30, then it would mean that patients need to not only satisfy MPR <80% constraint, but should also have at least one episode of medication lapse of 30 days.
The `NoMeasurementRecording` element is used to formulate queries related to timely measurement recording (criteria 2 and 3). The `ConsecutivelyUncontrolledMeasurements` element contains parameters to identify consecutively uncontrolled measurements. This criterion is an indication of time to achieve target, for example, if a patient has three or more consecutively high BPs over 180 days (criterion 4), then this indicates that the patient has not achieved goal BP within a reasonable length of time. The `MeasContraindicatingTherapy` element is needed to determine patients who have measurement(s) contraindicating therapy (criteria 7 and 8) – for example, treating patients with hypertension with thiazides may increase serum uric acid levels, and it is important to either change therapy or else prescribe medication to reduce this particular side-effect.

All classes of criteria in Figure 5.3 are derived from the `CommonCriteria` class which defines the basic properties of an audit report. The importance of using a run-in period (defined as an attribute in `CommonCriteria`) has been discussed previously in Section 3.2. The `afterClassifications` attribute is a means of specifying the classifications the patients need to have, hypertension and/or diabetes, for example. This has been specified as an optional attribute so that classification constraints need be specified only if needed; for instance, identifying patients with consecutively high BPs is important even if the patients do not have the formal ‘hypertension’ classification. Within the practices we were working with, classification details were fairly accurate in the patient EMR (i.e., almost all patients who have had consecutively high BPs over time also had a corresponding hypertension classification), but this may not necessarily be true across all practices.

The `CommonMeasurementCriteria` element contains attributes that are related to measurements. There is provision to specify two measurement threshold values along with corresponding comparison operators. The first is to be used for all measurement values (such as eGFR for example) as well as systolic BP. Currently, the only use of the second is to specify the diastolic BP threshold when measurement type is BLOOD_PRESSURE.

`ComparisonOperators` and `MeasuremetType` types have been defined as enumeration types within the model as shown in Figure 5.4. Attributes requiring a `DrugsOrTherapy` type in Figure 5.3 can be specified using generic drug names or by specifying a type of therapy such as ‘Antihypertensives’. The `Period` type has been specified using a start date and an end date.
Table 5.3 demonstrates the use of attributes, elements and enumerations of the Criteria Model by formulating the queries in Table 3.4 using the criteria elements shown in Figure 5.3. The *evaluationPeriod* and *runInPeriod* attributes are not shown as these two are the same across all the criteria.
**Table 5.3: Using elements from the Criteria Model to formulate criteria in Table 3.4**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Criteria Model Element</th>
<th>Values for Criteria Model Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>(per Table 5.3)</td>
<td>(per Figure 5.3)</td>
<td>(per Figure 5.3)</td>
</tr>
<tr>
<td>C1 MedicationLapse</td>
<td>afterClassifications = Hypertension</td>
<td>minLapseDuration = 30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>compareMinLapseUsingOp = GREATER_THAN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lapseInDrugsOrTherapy = Antihypertensives</td>
</tr>
<tr>
<td>C2 NoMeasurementRecording</td>
<td>afterClassifications = Hypertension</td>
<td>measurementType = BLOOD_PRESSURE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>minDaysWithoutMeasurement = 180</td>
</tr>
<tr>
<td>C3 NoMeasurementRecording</td>
<td>afterClassifications = Hypertension</td>
<td>measurementType = BLOOD_PRESSURE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>measurementThreshold1 = 160</td>
</tr>
<tr>
<td></td>
<td></td>
<td>measurementThreshold1Using = GREATER_THAN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>measurementThreshold2 = 100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>measurementThreshold2Using = GREATER_THAN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>minDaysWithoutMeasurement = 120</td>
</tr>
<tr>
<td>C4 ConsecutivelyUncontrolledMeasurements</td>
<td>afterClassifications = Hypertension</td>
<td>measurementType = BLOOD_PRESSURE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>measurementThreshold1 = 160</td>
</tr>
<tr>
<td></td>
<td></td>
<td>measurementThreshold1Using = GREATER_THAN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>measurementThreshold2 = 100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>measurementThreshold2Using = GREATER_THAN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>minConsecUncontrolledMeas = 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>uncontrolledMeasOverMinDays = 120</td>
</tr>
<tr>
<td>C5 MedicationLapse</td>
<td>afterClassifications = Hypertension and diabetes</td>
<td>minLapseDuration = 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lapseInDrugsOrTherapy = ACEis or ARBs</td>
</tr>
<tr>
<td>C6 MedicationLapse</td>
<td>afterClassifications = Hypertension and myocardial infarction</td>
<td>minLapseDuration = 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lapseInDrugsOrTherapy = Beta-blockers</td>
</tr>
<tr>
<td>C7 TherapyMeasurementOutOfTarget</td>
<td>afterClassifications = Hypertension and renal impairment</td>
<td>measurementType = eGFR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>measurementThreshold1 = 60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>measurementThreshold1Using = LESS_THAN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lapseInDrugsOrTherapy = ACEis or ARBs</td>
</tr>
<tr>
<td>C8 TherapyMeasurementOutOfTarget</td>
<td>afterClassifications = Hypertension</td>
<td>measurementType = URIC_ACID</td>
</tr>
<tr>
<td></td>
<td></td>
<td>measurementThreshold1 = 0.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>measurementThreshold1Using = GREATER_THAN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lapseInDrugsOrTherapy = Thiazides</td>
</tr>
</tbody>
</table>

### 5.5 The Computational Framework

ChronoMedIt’s computational engine has been developed using C# .NET 3.5 (simply referred to as C# hereafter) to process the audit criteria and identify the required patients. Microsoft SQL Server 2008 was used to store the patient data. An overview of the components of ChronoMedIt is shown in Figure 5.5.
The Controller module in Figure 5.5 communicates with the different modules to manage the overall process. The following sections describe the functionality of each of the other important components.

### 5.5.1 Creating Audit Criteria

The Reporting Criteria and the Reporting Criteria Template together define the content and the structure of the reporting criteria. The former is an XML document containing specifics of the criteria while the latter is an XML-Schema (XML-S) [261] document that specifies the structure of the criterion according to the Criteria Model discussed previously in Section 5.4. For instance, Figure 5.6 shows a snippet of the XML document that needs to be used to formulate the query ‘patients classified with hypertension and diabetes who have an ACEi/ARB MPR less than 80% with a lapse greater than 30 days’.
The XML tags in Figure 5.6 are specifying that I am interested in determining patients who have a medication lapse after a hypertension and a diabetes classification (the ‘^’ in the selected_problems tag is a separator indicating that patients should have both, hypertension and diabetes while ‘|’ indicates that patients should have at least one of the problems) who have a minimum lapse in ACEi/ARB medication for greater than 30 days and have an ACEi/ARB MPR less than 80%. Note that the term ‘less than 80%’ is used rather loosely, but as per Figure 5.6, this is strictly greater than 0% and less than 80%. Hence, in this case I am trying to assess adherence for patients under active ACEi and/or ARB management (with at least one ACEi/ARB prescription during the evaluation period or during the run-in period and extending coverage into the evaluation period).

This XML document is validated against the corresponding XML-S (note how ReportingCriteria.xsd has been specified in the case of Figure 5.6) by the XML-S Validator module. If validation was successful, the XML Criteria Parser module gets activated which then creates a
C# representation of the criteria from the XML document that can be processed by other C# modules.

5.5.2 The Drug and Classification Knowledge Base

The Drug and Classification Knowledge Base was modelled in OWL [233] using the Protégé-OWL [235] development environment and the previously discussed knowledge base (see Section 4.3.1) was used as a starting point. OWL was chosen as the modelling language for this work also, because of several reasons: (i) the resulting knowledge base (in fact, an ontology) is easily extensible and readily sharable, (ii) support for domain level modelling of therapeutic and classification knowledge (see Section 5.1 for domain level modelling requirements), and (iii) availability of free tools for (hierarchical) domain level modelling and knowledge visualisation (Protégé-OWL in this case).

The Drug Knowledge Base currently consists of antihypertensive, antidepressant, antidiabetic and statin drugs (based on generic names) and their corresponding drug classes. This knowledge base was created based on information extracted from JNC7 [6] as well as DrugDigest (http://www.drugdigest.org/ [accessed 20 January 2010]). During the data analysis stage, cases were identified where some drugs were prescribed using popular brand names instead of their generic names, and several such brand names have been included as well into the resulting knowledge base. The drug knowledge base consists of different drug classes such as ACEi/ARBs, diuretics and so on that are collectively categorised under the drug class ‘Antihypertensives’, leading to a drug class hierarchy where each drug belongs to at least one drug class.

The drug and classification information that was modelled using Protégé-OWL is shown in Figure 5.7. Note that this modelling exercise is different from the previous one shown in Figure 4.5 in that with the current approach, it is easier to build queries at an individual drug level as well; for example, a patient can get prescribed cilazapril as a single drug cilazapril (hydrochloride) or as a combination as cilazapril hydrochlorothiazide (i.e., a combination of ACEi and thiazide diuretic). With the current approach, the cilazapril class contains information about both drugs (note how two cilazapril drug instances can be observed in the Cilazapril_DrugClass in Figure 5.7) making it easier to build queries using this Cilazapril_DrugClass when wanting to create a query related to cilazapril. With the previous approach, drug classes could be easily included in queries (e.g., ACEi/ARB), however, building queries around individual drugs needed a greater effort. Also, the MedTech32_Specific_Variants
and BrandNames classes have been introduced to account for the different ways drug names had been entered during prescribing – for example, the generic drug name quinapril hydrochlorothiazide was entered as ‘quinapril;hydrochlorothiazide’ and ‘quinapril with hydrochlorothiazide’ in which case the variants were entered into MedTech32_Specific_Variants and OWL:sameAs links were created to the ‘quinapril hydrochlorothiazide’ individual to indicate that all these refer to the same drug. More rarely, brand names were used for prescribing, and for such cases, the brand names were created in BrandNames class with OWL:sameAs links to the generic name. An example of this would be prescribing ‘inhibace’ (a brand name) instead of ‘cilazapril’ (the generic name).

Figure 5.7: Two views of the drug and classification knowledge base. Part of the Antihypertensives class is expanded in the left view and the Antidepressants and Problems classes are expanded in the right view.

Drug Classes in Figure 5.7 is a high-level concept under which different drug classes have been created. Although the eight criteria are primarily hypertension related, several other drug classes (such as Antidepressants) have been created in the drug knowledge base (Figure 5.7) for illustration purposes to indicate how the drug ontology can be expanded to other domains. The Antihypertensives class further refines to different subclasses of antihypertensive medication. Similar to Drug Classes, the Problems class represents the different classifications of interest.
(modelling this knowledge was very similar to the previous effort discussed in Section 4.3.1) – in this case only diabetes, hypertension, renal impairment, myocardial infarction and depression have been included (depression has been included here to demonstrate how the Problems class is easily extensible to other clinical domain so that a criterion such as ‘patients with depression who have a lapse >30 days in their antidepressant therapy’ can be easily authored if required). Since Read Clinical Codes are used in NZ to classify patients, only the relevant Read Codes have been created in the Read_Codes class which is one of the ClassificationSchemes. To indicate that a particular Read Code denotes a condition, the corresponding condition classes (Diabetes and Renal_Impairement, for example, in the case of Read Code C104.11 which indicates that a patient has diabetes and hence developed renal impairment) is also made a parent class of that Read Code (i.e., multiple inheritance). The grayish number within parenthesis indicates the number of OWL individuals [233] in each class; for example, five Read Codes are associated with hypertension (represented by the Hypertension class in the ontology). If terms from a new coding scheme (ICD for example) are to be introduced, this new scheme can be created as a child class of ClassificationSchemes and the different problems can be made parents of the corresponding classification codes (i.e., the specific instances of the new coding scheme). The end result of this modelling effort would be the problem types having classification codes from different coding schemes associated with them. This type of approach promotes consistency of higher/domain level concepts (e.g., ‘hypertension’) while effectively abstracting the lower level system/implementation specific details (e.g., Read Code G2.00 with associated Read Term ‘hypertensive disease’ to indicate ‘hypertension’).

5.5.3 ChronoMedIt’s Database Structure and Parsing BPs and Lab Test Results

The ‘Patient Data’ in Figure 5.5 represents a Microsoft SQL Server 2008 database that contains all the patient data. SQL Server was chosen as the preferred database compared to the previous choice of Microsoft Access primarily because of the stored procedure capabilities. These procedures could be used to process data on the server itself whereas with Microsoft Access, any processing would need to be done at the client side, or within the database using sequential processing via VBA modules.

ChronoMedIt uses five main tables to store patient data; Figure 5.8 shows the table definitions that have been used. int, date and float are standard data types, and varchar represents a string type with the number in parenthesis showing the maximum length of the string.
Although the data needs to be in the format required by the tables shown in Figure 5.8 for various queries to be processed, this is not the format the raw data is in when a dataset is extracted from the commercial PMS system. I worked with the MedTech32 system (the most widely used PMS system by NZ GPs which has about 80% market share in NZ) and Section 3.3.1 briefly discussed some of the data cleaning/pre-processing that was required. Towards this stage of my PhD, my supervisor and I were looking at working with multiple GP practices, and the previously used (rather manual and VBA based) processing needed to be more automated. Importing most of the extracted data directly to the database was fairly straightforward, however, processing the BPs from notes, and extracting lab test results from the ‘Provider Inbox’ needed to be more automated. The notes and ‘Provider Inbox’ are ‘blob’ (Binary Large OBject) fields within the database meaning that this is essentially an array of bytes stored in the database. Blobs can be used to store formatted documents (among other uses) and do not have a single specific data type; they may contain control characters, and therefore cannot be readily processed. Note that the particular data cleaning issues discussed below are rather MedTech32 specific, yet warrant some discussion as it was an important part of the data cleaning effort.

Within MedTech32, there are two areas where GPs can enter notes into the patient record during consultations – the Subjective notes section and the Objective notes section (see Figure 5.9).
Often clinicians use the notes areas to record BPs as well, however, as mentioned in Section 3.3.1 it is important for the BPs to be entered following the convention `\bp`. If this is not followed, the system will not automatically extract these BPs into the database table that contains BP measurements (which can also be populated using a form specifically designed to enter BPs). Analysing these notes fields showed that there are many instances where the busy GPs would simply enter something like ‘bp 160/100’ instead of following convention. This means that if the BPs extracted only from the BPs table of the PMS, quite a few measurements will be missed out. As previously mentioned, the notes are stored as a blob field in the database and contains many control characters; for example, the notes for a hypothetical patient may resemble those shown in Figure 5.16 after exporting the blob into a text file. The first row shows the different fields within the notes.

With respect to lab test results, the NZ system is such that the results get sent to the GPs electronically, which in the case of MedTech32 get stored in the ‘Provider Inbox’. A GP can easily look at these values by opening the corresponding record, however, if we need to analyse say uric acid levels for all patients in the practice, there is no way to extract this information at a
practice level. Shown in Figure 5.17 is what a test result would look like within the PMS (the data shown is from the training dataset that gets shipped out with a new MedTech32 installation). 

![Figure 5.11: Viewing lab test results within MedTech32](image1)

Using MedTech32’s built-in SQL query tool, one can extract the Provider Inbox data to a text file, but assuming Charlene Abbott’s (i.e., the selected patient in Figure 5.17) MMID is ABCDEF, the resulting text would look like that shown in Figure 5.12. 

![Figure 5.12: A ‘blob’ representation of a lab test result](image2)

As a solution for these two issues, I developed a C# based tool to extract BPs from notes, and lab test results from Provider Inbox. The tool uses C# regular expressions to extract values of required type and is dependent on having two text files on the local disk: the file containing subjective and objective notes; and the lab test results (both files can be created using the built-in SQL query tool within MedTech32). I was primarily interested in extracting only albumin-creatinine ratio (ACR), Uric Acid, eGFR, HbA1C and Microalbumin and as such the tool was programmed to extract only these labs (more lab tests than required for the eight criteria to facilitate other ongoing research clinicians were interested in). However, extending the tool to extract other measurements is possible and was to be done if/when necessary.
The main interface of this is shown in Figure 5.13.

![Figure 5.13: A tool to extract BPs from clinical notes and lab test results from MedTech32’s Provider Inbox](image)

The tool is trivial to use and selecting Options → Parse BPs brings up the dialog box to select the file with subjective/objective notes. Then clicking ‘Parse BPs’ creates a new file ‘CLEANED_BPs’ in the same folder as the subjective/objective notes file. This file is well formatted and can be directly imported into the patient database.

‘Parsing labs’ is similar; Selecting Options → Parse Labs and then selecting the file with lab inbox data will parse the file for the required lab type and create a new file ‘CLEANED__’ followed by the name of the selected lab test. Another option is to select ‘All Labs’ which will parse the lab inbox file to extract all lab test results (only for the set of lab tests previously mentioned). Figure 5.14 shows an example of parsing for ‘All Labs’.
The tool also allows the data to be sorted by the different columns if required. The information is editable, so if there is a logical error, such as the GP inadvertently entering a BP value of 1309/5 instead of 130/95, the tool will show a systolic/diastolic value of 1309/5. The user can manually correct such issues and then do Options → Export Grid Data to export the updated file into a new text file which can then be imported into the patient database.

5.5.4 Implementation of the Criteria Model

All the methods required to implement the Criteria Model are defined in the ICriteriaManager interface (see Figure 5.15) that is defined within the Criteria Processor module. This ICriteriaManager interface is the main accessor method to retrieve patients who satisfy a given criterion. The GetPatientsSatisfyingCriterion(ICriterion criterion) is the main method which accepts an instance of a criterion that implements the ICriterion interface. Although not shown in Figure 5.3, all the criteria elements (MedicationLapse, NoMeasurementRecording for example) implement the ICriterion interface which is a very simple interface that specifies a few criterion independent methods. The GetPatientsSatisfyingCriterion method returns an array of standard .NET DataTables and the number of tables the array contains varies depending on the criterion of interest. If the criterion was MedicationLapse with only a lapse or MPR specified after the required classifications, then the resulting array will contain two DataTables, the first one a list of patients with corresponding lapse or MPR details and the second table containing details related to the patient classifications. If both lapse and MPR were specified with the corresponding classifications, then the resulting array will contain three DataTables – the first
will be lapse details, the second will be MPR details and the third will be the classification details. Returning an array of standard .NET DataTables in this manner has the advantage of having query results in a consistent manner irrespective of the criterion. The resulting DataTables can then be used to present the data to a clinician in a meaningful manner – for example, the underlying data source for the criteria reports discussed in Section 5.5.5 are these DataTables. On the other hand, if criteria specific data structures were used, an end user will need to become familiar with these data structures requiring a greater development effort.

The ICriteriaManager interface in Figure 5.15 also specifies around 25 different carefully identified helper methods (only two are shown in the figure) where the required parameters need to be passed into the method to query for patients. These helper methods are based on variations of the Criteria Model elements – for example, there are three helper methods (i.e., variations) for the MedicationLapse criterion – lapse only, MPR only and lapse and MPR (out of which only the first variant is shown in Figure 5.15).

The SQLCriteriaManager implementation is primarily based on SQL Server stored procedures and functions and is the main functional unit of the SQL Criteria Processor module in Figure 5.5. The SequentialCriteriaManager has a C# based procedural implementation and forms the main functional unit of the Sequential Criteria Processor module in Figure 5.5. As such, these two are independent implementations of the Criteria Model that implement the methods specified in ICriteriaManager. These two implementations have been used as the basis for verification of the
framework via random testing – see next chapter for details. This type of multiple implementations reflects the Union design pattern in software engineering and is based on polymorphism concepts of object oriented design.

5.5.5 Framework Outputs

The two key outputs of ChronoMedIt are criteria reports and prescription charts. Criteria reports are .pdf documents that are generated using the .NET ReportViewer control by the Reporting Client module shown in Figure 5.5. The Reporting Client communicates with the Criteria Processor and all reports are dependent on DataTable[] returned by the various ICriteriaManager methods. These reports contain lists of patients with the relevant patient details indicating how a given patient satisfies the criterion constraints. This subsection shows the different kinds of reports ChronoMedIt can produce as relevant to the four classes of audit criteria along with the corresponding XML document. For each criterion, an exemplar case is demonstrated using a novel prescription visualisation tool I have developed. It needs to be noted that for all the cases discussed here, at least a hypertension classification has been specified (using the <selected_problems> XML tag), but if needed, all queries can be formulated and the corresponding report generated without specifying a classification requirement as well.

5.5.5.1 Adherence/Persistence to Indicated Medication

Criteria that fall into this category can be specified in terms of lapses (persistence) and/or MPR (adherence). For example, Figure 5.6 showed the XML for the query ‘patients classified with hypertension and diabetes who have an MPR <80% with a lapse >30 days for ACEi/ARB medication for the evaluation period 01-May-08 to 30-Apr-09’. Details for a selected patient from the resulting report for this query are shown in Figure 5.16. It should be noted that this report includes lapse, MPR and classification details that need to be satisfied, but the query can be formulated using any combination of these as well (except ‘classifications only’).
Figure 5.16: Details for a patient who satisfies criterion ‘ACEi/ARB MPR less than 80% with lapse greater than 30 days after classified with hypertension and diabetes’

The top of the report displays the explicit reporting criteria in a more human friendly format (than the original XML document that was used to specify criteria details). The Patient Id is the practice-specific patient identifier the researchers were given, as per ethics protocol. Patient MPR for selected drug class is then shown followed by details of the different lapses that satisfy the inclusion criteria. The corresponding patient diagnoses details – i.e., the Read Terms (non-insulin dependent diabetes mellitus and hypertensive disease in this case), the Read Codes and the classification dates are also shown.

With most PMSs, the presentation of prescribed medication for a patient is often in the form of a table where the prescriptions are ordered chronologically. It is a challenge to identify patients with prescribing lapses with this form of presentation. As a possible solution, I developed a C# based graphical tool (this component is represented by Prescribing Charts in Figure 5.5) that can assist a clinician to easily visualise patient prescribing patterns and also to present some feedback to the patient if required in a more intuitive and visually appealing manner. This was developed using some freely available software written in Visual Basic .NET 2.0 as a starting point [262]. A clinician can enter the patient identifier into the framework which will produce...
the prescribing plot for that patient. Development of this visualisation tool was driven by the
need to have a more interactive tool (e.g., filtering by drug class) compared to the previously
presented Excel based visualisation tool.

The visualisation tool also helps in visually validating the cases identified by ChronoMedIt as
being on suboptimal therapy. For example, the prescribing chart in Figure 5.17 corresponds to
the prescriptions issued to the patient identified in Figure 5.16 as being non-adherent.

![Figure 5.17: A prescribing chart for a selected patient](image)

The chart indicates the different drugs the patient has been prescribed using a different colour
to represent each drug to aid in following a single drug on the temporal axis. The drop-down
option on the top-left provides the various drug classes the prescribed drugs belongs to (based
on the Protégé ontology), effectively implementing a drug class filter (in Figure 5.17 All Drug
Classes has been selected, therefore all prescriptions are shown). The prescribing chart also
shows the selected evaluation period (01-June-07 to 31-May-08 in this case). Shown in Figure
5.18 is a filtered version of Figure 5.17, which shows only the prescriptions with an ACEi (per
Figure 5.17, this patient has no ARB prescriptions, so the MPR/lapses are for ACEi only). The
figure also shows how hovering the mouse over a prescription bar brings up a tooltip that
includes prescription duration and the different drug classes it belongs to.
It should be noted that in Figure 5.18 there are three lapses in ACEis, although in Figure 5.16, only two have been reported. The report in Figure 5.16 shows ACEi/ARB MPR is 67.4% but if we do a simple MPR calculation based on these two lapses, that is \(\frac{365 - (33+75)}{365}\), we get 70.4%. This is not a discrepancy in results, but is due to the fact that there is a lapse of 11 days from 07-Dec-07 to 18-Dec-07 (this can be seen if we hover the mouse over the relevant prescriptions) which is not shown in Figure 5.16, as the query was only for lapses greater than 30 days. Now if we perform the MPR calculation, that is \(\frac{365 - (33+75+11)}{365}\), then we get the same 67.4% which is in fact what is reported in Figure 5.16.

5.5.5.2 Timely Measurement Recording

Figure 5.19 shows the XML needed to formulate a query to determine patients with hypertension who have a lapse in their BP measurement for at least 120 days after having a high BP measurement (i.e., \(BP \geq 160/100\) mmHg). The \(<after_measurement_thresholds>\) XML tag is an optional tag so that a query such as ‘patients with no BP measurement for 180 days’ (per C2 in Table 3.4) can be easily formulated.
Figure 5.19: The XML document to determine patients with hypertension who have a lapse in BP measurement for over 120 days after a high BP measurement

Details for a selected patient from the resulting report are shown in Figure 5.20.

Figure 5.20: Details for a patient with no BP for ≥120 days after a high BP measurement
The *Measurement Charts* component in Figure 5.5 is used to display measurements such as BPs, eGFRs and so on (the drop down in the measurements chart in Figure 5.21 can be used to select between the different measurements the patient has). This was developed as an extension to the prescription charts discussed before so that uncontrolled measurements (such as high BPs) can be easily identified along the same temporal axis as medication. Visualising such measurements on the same axis can be used to monitor patient improvements against therapy. Figure 5.21 shows a plot where the prescriptions and BP measurements are displayed on the same temporal scale.

![Figure 5.21: A prescribing and a BP measurement plot](image)

Based on the graph, it is reasonable to interpret that at the beginning, the patient’s diastolic BP lowered from 130 mmHg to 110 mmHg while taking the prescribed medication, but BP increased to 180/120 mmHg after having a significant medication lapse. Again BP has started decreasing until it reached 110/80 mmHg before increasing again to 140/100 mmHg, the increase occurring possibly due to the medication lapse.

### 5.5.5.3 Time to Achieve Target

This criterion class can be used to formulate queries to determine patients who have consecutively uncontrolled measurements, for example, three or more consecutively high BP
measurements over 120 days or more. In the XML shown in Figure 5.22, ‘high’ has been specified as 160/100 mmHg.

```xml
<?xml version="1.0" encoding="utf-8"?>
<Report xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:noNamespaceSchemaLocation="ReportCriteria.xsd">
  <evaluation_period>
    <start_date>2008-05-01</start_date>
    <end_date>2009-04-30</end_date>
  </evaluation_period>
  <run_in_start>2007-11-01</run_in_start>
  <consecutively_uncontrolled>
    <min_consec_uncontrolled_meas>3</min_consec_uncontrolled_meas>
    <measurement_type>
      <blood_pressure>true</blood_pressure>
    </measurement_type>
    <meas_over_min_days>120</meas_over_min_days>
    <after_measurement_thresholds>
      <min_value>
        <min_threshold_value>160</min_threshold_value>
        <min_threshold_value>100</min_threshold_value>
        <compare_min_threshold_using>GREATER_THAN_OR_EQUAL_TO</compare_min_threshold_using>
        <true_value>true</true_value>
      </min_value>
    </after_measurement_thresholds>
    <selected_problems>Hypertension</selected_problems>
  </consecutively_uncontrolled>
</Report>
```

Figure 5.22: The XML document to determine patients with hypertension who have three or more consecutively high BP measurements over 120 days or more

A snippet from the corresponding report for a selected patient is shown in Figure 5.23. Note how four consecutively high BPs have been grouped under ‘1’. This indicates that for this patient there was only one sequence of uncontrolled BP measurements. If the patient had this sequence, followed by a ‘controlled’ BP and then three or more consecutively high BPs, then those BPs would be shown under a new sequence grouped under ‘2’.
The prescription-BP chart in Figure 5.24 shows an alternate view of the different BPs the patient has had. Similar to the patient in Figure 5.21, the lapse in medication may have been the reason for the increased BP of 200/110 mmHg that has been observed.
5.5.5.4 Measurement Contraindicating Therapy

Figure 5.25 shows the XML required to formulate a query belonging to this criterion class. The specific example shows the tags needed to formulate the query ‘patients with hypertension who have a high uric acid level (defined as >0.42 mmol/l) while on thiazide diuretics’.

```xml
<?xml version="1.0" encoding="utf-8"?>
<Report xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xmlns:xsd="http://www.w3.org/2001/XMLSchema"
     xmlns:ns="urn:xol-initiative:report_criteria">
  <evaluation_period>
    <start_date>2008-05-01</start_date>
    <end_date>2009-04-30</end_date>
    <run_in_start>2007-11-01</run_in_start>
  </evaluation_period>
  <therapy_measurement_out_of_target>
    <measurement_type>
      <uri>http://example.org/TherapyType</uri>
    </measurement_type>
    <selected_therapy>
      <selected_drug_class>Thiazide_Diuretics</selected_drug_class>
    </selected_therapy>
    <after_measurement_threshold>
      <min_meas_value>
        <min_threshold_value>0.42</min_threshold_value>
        <compare_min_threshold_using>
          <GREATER_THAN>true</GREATER_THAN>
        </compare_min_threshold_using>
      </min_meas_value>
    </after_measurement_threshold>
    <selected_problems>Hypertension</selected_problems>
  </therapy_measurement_out_of_target>
</Report>
```

Figure 5.25: The XML document to determine hypertensive patients who have a high uric acid value while on thiazide diuretics

Figure 5.26 shows part of the resulting report for a selected patient while Figure 5.27 shows the prescription-uric acid plot for this patient.
In this particular instance, it appears that the uric acid levels have slightly increased over time and if this criterion was used to actively identify and manage patients, this patient would have got prescribed allopurinol in an attempt to control the uric acid levels and also received close medical follow-up if needed.
5.5.6 The Graphical User Interface

Figure 5.28 shows the main GUI of ChronoMedIt. The interface is divided into four main areas as indicated in Figure 5.28. Area (1) is used to show the different criteria classes a user can select. Once a selection is made, the corresponding panel with the respective controls gets activated and shown in area (3). Area (2) is a ‘header’ area common to all criteria classes where the evaluation period and the run-in period can be specified. Area (4) is a read-only area which automatically gets populated as different selections are made within a criterion. For instance, in Figure 5.28, MPR and lapse constrains are both selected, and if the lapse duration checkbox is unchecked, area (4) will automatically get updated to reflect this change. Once the required criterion has been formulated, pressing the ‘Generate Report’ button will create the corresponding report for the specified criterion (the various reports were discussed previously in Section 5.5.5).

When the user interface first opens, it defaults to ‘Lapse in Therapy’ where the default criterion gets set to ‘patients with antihypertensive MPR greater than 0% and less than 80%’. The reporting period (i.e., the evaluation period) defaults to the past year with the end of the evaluation period being that day; the run-in period is given a default value of 6 months.
The classification and drug/therapy details can be modified by clicking the corresponding ‘Select’ button in the main interface. The classification selector that opens up is shown in Figure 5.29.

![Classification Selector](image)

Figure 5.29: The classification selector

The ‘Selected Classifications’ text automatically gets updated as different checkboxes are checked and unchecked. As shown in Figure 5.29, specific codes within a problem can be selected as well which are joined using ORs while domain level problems are joined using AND, as in ‘Hypertension and Diabetes’ for example. The tree view shown in Figure 5.29 has been implemented as a generic component so that checking a parent node will automatically check all children (recursively), and checking all children will automatically check the corresponding parent. Similarly, unchecking a child node when all children are checked (hence the parent will also checked) will automatically uncheck the parent node.

Figure 5.30 shows how the different drug classes can be selected. Similar to the classification selector, the drug ontology selector has been implemented using a ‘tree’ control, however, the selection of drugs/drug classes is always joined using ORs. How ‘thiazides or ACEis or ARBs’ can be specified is shown in Figure 5.30. Although not shown here, any individual drugs within the same/different drug classes (e.g., cilazapril OR quinapril) can also be selected if needed.
Once the required criterion has been formulated using the main user interface, the criterion can then be saved as an XML document, as shown in Figure 5.31. The generated XML document tags are similar to the ones discussed this far, specifically in Section 5.5.5, and conform to the same XML-Schema document (i.e., the .xsd file) mentioned in Section 5.5.1. The only exception to the tags seen previously is an optional $\texttt{<notes>}$ tag which is a text field that can be used to store any details related to the criterion. We envisage this interface being used (possibly by a practice manager) to formulate the ‘query templates’ that need to be run on a regular (quarterly) basis and therefore having this type of a feature to save query details was seen to be useful.
A saved XML document/query template can also be imported into the system by going into the ‘Open’ menu within the ‘File’ menu which brings up the dialog shown in Figure 5.32. Upon selecting a valid file, the criterion described within the XML will be loaded into the main user interface.

The user interface also has the option (Figure 5.33) to generate prescribing charts.
The tool that generates the prescribing charts needs the patient identifier, the evaluation period and the start and end dates for the chart (Figure 5.34). The default is set to be the evaluation period shown within the main header panel (Figure 5.33) with the plot start set to a year prior to evaluation period start and the plot end set to six months after end of the evaluation period. This extra duration before and after the evaluation period is needed to attain a better view of a patient’s therapy so that the prescriptions running-into the evaluation period and the end dates of the prescriptions issued just prior to end of the evaluation period can be visualised.

Once the parameters have been specified, the ‘Plot’ button can be pressed to bring up the corresponding prescribing charts discussed previously in Section 5.5.5. Figure 5.28 showed how a ‘Lapse in Therapy’ criterion could be formulated using the main user interface. Similarly, clicking on the different criterion classes in Area (1) activates the corresponding criterion.
formulation panel which is then displayed in Area (3). For instance, the panel that gets activated when ‘Lapse in Observations’ is selected is shown in Figure 5.35.

![Figure 5.35: Formulating a ‘Lapse in Observations’ criterion](image)

### 5.5.7 Criteria Processing with SQL or Sequential Implementation and Verification of Results

The nature of a SQL query is such that it will almost always return some result, unless there was an obvious syntax error (and is almost certain to return a result via a visual query builder). Therefore, it is necessary to verify that the SQL based implementation produces the correct results. Software verification usually refers to evaluating a software system by determining whether the product of a given development phase satisfies the requirements established before the start of that phase and is about building the product correctly [246].

The independent SQL Criteria Processor and the Sequential Criteria Processor implementations mentioned in Figure 5.5 were used as a solution for this. As mentioned previously, the SQL Criteria Processor uses a combination of SQL Server stored procedures and functions to determine patient cohorts that satisfy a criterion while the Sequential Criteria Processor uses a procedural based approach. For example, the C# based procedural implementation to determine medication lapses uses a discrete-event simulation model to determine the patients of interest. The related computation unit retrieves the knowledge on the required criteria details (lapse and/or MPR thresholds, type of medication and so on) by communicating with the C#
XML Criteria Parser module. Once the criterion is formulated, the pseudo-code to identifying gaps is:

- For each patient prescription, create two prescription events, one a start-event and the other an end-event denoting the start and end dates of prescription coverage.
- Order the prescription events by (date) ascending order.
- Loop through the prescription events; incrementing a counter by 1 if it is a start event and decrementing the same counter by 1 if it is an end event. If counter equals zero, then it is a lapse in medication.
  - If we are interested in lapse durations (i.e., as specified in the XML document) and the lapse satisfies the required temporal constraints and is after the required patient classification(s) (per Figure 5.1), then it becomes a valid lapse that we need to report.
  - If MPR criteria has been specified in the XML, then add the lapse duration (or the portion of the lapse duration per Figure 5.1) to TotGap so that we can calculate MPR later.
  - If MPR is required, calculate MPR using equations (3) and (4) now that we know the value of TotGap.
  - Consider the special cases (per Figure 5.1) where there may be no knowledge of any prescriptions, and therefore will not be picked up above.
  - Report on satisfied patient details, tailoring the output to the specifics of MPR and/or lapse criteria required.

The essence of this verification process is the comparing of the outputs from the two independent implementation modules and ensuring that the results returned by the two different mechanisms are the same. The Verifier Results (which reflects a system output) shown in Figure 5.5 is representative of the output from the SQL-Sequential Verifier module and contains details of any discrepancies that may have occurred if verification failed. A run with failed verification indicates an error somewhere, either in the SQL or sequential implementation, or a patient case not accounted for during development; hence each time verification failed, the issues behind the discrepancies in the results were identified and the relevant implementation was rectified. This was carried out as an iterative process until both implementations resulted in the same number of patients for a given criterion and satisfies the first framework verification requirement per Section 5.2. In order to satisfy the second verification requirement, I wrote a number of test cases to perform boundary value analysis, equivalence class testing and all-pairs testing for various test cases that I developed to cover the breadth of possibilities derived from the Criteria Model (per Figure 5.3). This provides the basis for the next chapter on framework testing.
Chapter 6
Framework Testing

Thus far, I have been discussing the different components of the computational framework and the overall system architecture. As with any software system, it is important to ensure that the framework’s final implementation is in concordance with its initial specification and intended use. In the software development life cycle, this type of software quality assurance is typically performed using software testing mechanisms and this chapter is devoted to discussing the different software testing techniques I have used to test the framework. A broad overview of some of the key software testing concepts (as relevant to this work) is first presented in an attempt to establish a common ground and explain the terminology. This is followed by details of the various test cases that have been developed to test the individual classes of audit criteria. I then discuss how the framework has been tested with a large number of randomly generated test cases to ensure reliability along with some background related to random testing techniques. Software testing on its own is a very broad field consisting of a vast number of important themes and the reader is referred to [245, 263-266] for a detailed presentation of different testing techniques.

6.1 An Overview of Software Testing Concepts

6.1.1 A Basic Introduction

Here I summarise some important testing concepts, as relevant to this work, using [245] as the primary source of reference.
Software testing usually refers to evaluating the features of a program against a given set of system requirements (i.e., the specification) with the intention of finding errors in order to verify that the program functions properly. Given a program and its specification, we can represent the logical relationship between these two behaviours using the Venn diagram shown in Figure 6.1.

![Figure 6.1: Specified and implemented program behaviour](image)

Of all the possible program behaviours, the specified ones are in the circle labelled S and all those behaviours that have been actually programmed are in P. This diagram shows how certain behaviours may be specified but not implemented (S ∩ ~P section) and, conversely, how certain behaviours may be implemented, although not specified (~S ∩ P section). The intersection of the two types of behaviour (i.e., the S ∩ P section) is the ‘correct’ portion, that is, program behaviour that has been both specified and implemented. A practical view of software testing is about determining the extent of program behaviour that is both specified and implemented.

The model shown in Figure 6.1 can be extended by adding the testing space to the set of behaviours, as shown in Figure 6.2 (where the testing space is denoted by ‘T’).
Considering the relationships among the sets S, P and T, there may be specified behaviour that is not tested (regions 2 and 5), specified behaviour that is tested (regions 1 and 4) and tested behaviour that is not specified (regions 3 and 7). Similarly, there can be programmed behaviour that is not tested (regions 2 and 6), programmed behaviour that is tested (regions 1 and 3) and tested behaviour that is not implemented (regions 4 and 7). Each of these individual regions is important as each is an indication of non-existing test cases for specified and/or programmed behaviour as well as testing non-specified and/or not-implemented program functionality. The goal of testing is to make the region where all three sets intersect (region 1) as large as possible.

Two fundamental approaches are typically used to identify the particular test cases denoted by the set T in Figure 6.2 – functional and structural testing. Functional testing (also commonly referred to as black box testing) is based on the view that any program can be considered to be a function that maps values from its input domain to values in its output range and the only information used is the specification of the software [245]. Structural testing (also called white box testing) uses the system implementation to identify the test cases which is its main difference from functional testing. In terms of the Venn diagram in Figure 6.2, functional testing is denoted by \( S \cap P \) section (regions 1 and 4) while structural testing is denoted by \( P \cap T \) section (regions 1 and 3). Whether functional or structural testing is better has been a debate that has continued for decades with strong views for either choice, but both types of testing have their own merits. Therefore, in order to make region 1 as large as possible, it is important to develop both types of test cases [245]. An important aspect of this work is to ensure that the different components representing the audit criteria satisfy their initial specification, irrespective of how the criteria have been implemented. As such, for the purposes of this thesis, the primary focus has been on functional testing.
Software verification and validation are two important concepts of software testing [266]. Verification usually refers to evaluating a software system by determining whether the product of a given development phase satisfies the requirements established before the start of that phase and is about building the product correctly. On the other hand, validation refers to confirming that a product meets its intended use, usually to confirm that a product meets its customer expectations. Verification activities are typically performed during the lifecycle of a project while validation is usually performed on the entire system by actually running the system in its real environment, normally towards the end of a project [266]. The ChronoMedIt framework is not a final product per se, and from a practical perspective, developing a complete system and performing validation with clinicians and/or practice managers as potential ‘customers’ of the system is not feasible; hence only verification of the framework has been carried out using the testing techniques presented in this chapter.

6.1.2 Types of Functional Testing

This section discusses the four key functional testing mechanisms: boundary value analysis, equivalence class testing, all-pairs testing and special value testing. The main focus has been on input domain testing since the relevant values from the input space can be systematically identified given the nature of the criteria to be tested; however, performing output domain testing was also a possibility.

6.1.2.1 Boundary Value Analysis

Boundary value analysis (BVA) focusses on the boundaries of the input space to identify the test cases. The rationale behind BVA is that errors tend to occur near the extreme/boundary values of an input variable [245, 266] and is based on the important assumption that is known as the single fault assumption in reliability theory – that is, a failure is only rarely the result of the simultaneous occurrence of two (or more) faults. The basic idea in BVA is to use input variable values at their minimum (min), just above the minimum (min+), a nominal value (nom), just below their maximum (max-) and their maximum (max) while holding all other variable values at their nominal value [245]. For example, considering the function F(x1, x2) with x1 having the range [a, b] and x2 having the range [c, d], then the test cases for this function are those shown in Figure 6.3.
Robustness testing is a simple extension to BVA by adding two extra test cases (Figure 6.4) - a value slightly exceeding the maximum ($max+$), and a value slightly less than the minimum ($min-$). The importance of robustness testing is that it forces attention on handling exceptions in a controlled manner. Therefore, whenever the boundary conditions needed to be tested, robustness testing has been chosen over BVA for all single fault testing purposes. Note that the term BVA is often used even when referring to testing with invalid values (instead of the more specific term ‘robustness testing’), and as such the terms BVA and robustness testing have been used interchangeably herein.

Special value testing is another widely used BVA technique, but it is also the least uniform since the test cases are based on domain knowledge with ‘best engineering judgement’ rather than a
systematic approach [245]. Some special value test cases that have been developed for testing ChronoMedIt are discussed in Section 6.3.

6.1.2.2 Equivalence Class Testing

Equivalence class testing refers to identifying the different partitions of a set, where a partition refers to a collection of mutually disjoint subsets, and the union of the partitions forms the entire set [245, 266]. The idea here is to identify the set of test cases by using one element from each equivalence class. The use of equivalence class testing has two important features: a sense of complete testing since the entire set is represented, and non-redundancy caused by disjointedness of the partitions. This type of testing also echoes two features of BVA – normal vs. robustness testing (i.e., the test cases shown in Figure 6.3 and Figure 6.4 respectively) as well as the single vs. multiple fault assumption (also referred to as weak vs. strong testing respectively). This leads to four forms of equivalence class testing, namely, weak-normal equivalence class testing, strong-normal equivalence class testing, weak-robust equivalence class testing and strong-robust equivalence class testing [245].

6.1.2.3 All-Pairs Testing

All-pairs testing, also referred to as ‘pairwise testing’ [266], is a technique where the test cases are identified by using the distinct pairs of all input parameters to the system [245, 266]. The reasoning behind this is that most of the bugs in software systems occur due to single faults rather than multiple faults occurring simultaneously. Wallace and Kuhn [267] in their studies on defects in software-controlled medical systems reported that 98% of the defects were due to the interactions of pairs of variables, and that defects involving interactions between three or more parameters are progressively less common. They also reported that testing for combinations of three or more faults becomes progressively more expensive, which has as its limit the exhaustive testing of all possible input combinations. The number of test cases for all-pairs testing for \( n \) variables can be calculated using the mathematical formula for combinations:

\[
\binom{n}{2} = \frac{n!}{2!(n-2)!}
\]  

(6.1)

Therefore, whenever strong testing was required, in an attempt to reduce the number of test cases, all-pairs testing has been carried out instead of exhaustive testing which would have resulted in \( 2^n \) test cases.
6.2 Methods

The general notion around testing the framework is to (i) develop the test cases that adequately cover the audit criteria the framework implements; (ii) implement the test cases using a suitable software testing framework; (iii) execute the test cases against the sequential based implementation (Section 5.5.4), and (iv) analyse results of the test runs and iteratively rectify any software issues (aka ‘software bugs’) until all the tests are passed. Once the sequential implementation has been tested and verified with reasonable confidence, the next phase of testing involves developing a test harness where a large number of test cases can be generated and the results of the SQL based implementation can be compared against the results of the sequential based implementation.

The test cases that have been developed to test the sequential implementation are discussed in Section 6.3. The framework has been developed in C# .NET using Visual Studio 2008, and therefore the natural and the most logical testing environment of choice was MSTest [194] – a software unit testing framework for .NET that is closely integrated with the Visual Studio development environment. Using this testing framework, the test cases were created as memory based test cases in C#, and then executed. If any of the tests failed, then the related, erroneous lines of code in the sequential implementation were analysed and corrected. This process was performed iteratively until all the test cases passed.

As mentioned previously, it requires a great deal of effort to test each and every possible scenario related to an audit criterion by writing a corresponding test case to represent the scenario. Random testing provides a potential solution to this problem and Section 6.4 presents the related details.

6.3 Identification of Test Cases

This section discusses the various test cases used to test the different classes of audit criteria that have been established in previous chapters. All test cases have been developed with respect to a six-month evaluation period from 01-Jul-08 to 31-Dec-08 and a six-month run-in period from 01-Jan-08 to 30-Jun-08.

Note that an interval is represented using |——|, and in the context of persistence of medication represents a lapse duration while an interval in the context of compelling indications represents
the duration of available medication (provided that it was consumed as directed). A dark dot (●) represents a measurement (i.e., an event/point in time rather than an interval).

6.3.1 Test Cases for Persistence of Medication and Timely Measurement Recording

The first audit criterion involves identifying periods of medication lapse where the lapse duration and location need to meet certain criteria constraints. For this case, the weak-robust equivalence classes (weak testing for single faults, and robustness testing for invalid values) that need to be considered when the diagnosis date is prior to evaluation period are shown in Figure 6.5.

![Figure 6.5: Weak-robust equivalence classes for medication lapse with diagnosis date prior to evaluation period (EP). Scenarios 1-7 represent lapse periods.](image)

There are seven distinct regions a lapse can occur where each region has a before, after or during type relationship to the diagnosis date, and a before, overlapping, during or after type relationship to the evaluation period. We have carefully constructed these lapse durations by creating prescriptions around the required lapse duration so that no other lapse occurs, and also so that any two lapse durations can occur concurrently (required for all-pairs testing). For example, regions 3 and 4 do not overlap, because if they did, lapses in region 3 and 4 that occur concurrently will not be logically meaningful.

We are trying to determine medication lapses of at least 30 days where the lapse overlaps with the evaluation period, and the lapse starts after a particular diagnosis. Therefore, all test cases have been designed with a nominal 45-day lapse duration. Table 6.1 shows the input
parameters and expected output details for the test cases shown in Figure 6.5. For all test cases, a Case ID has been assigned where the numeric value reflects the corresponding scenario in Figure 6.5.

Table 6.1: Test cases for weak-robust equivalence class testing for medication lapse with diagnosis prior to evaluation period

<table>
<thead>
<tr>
<th>Case ID</th>
<th>Diagnosis Date</th>
<th>Lapse Period</th>
<th>Is Valid Lapse?</th>
<th>Valid Lapse Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>WR1</td>
<td>01-Mar-08</td>
<td>04-Dec-07 to 18-Jan-08 (45 days)</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>WR2</td>
<td>01-Mar-08</td>
<td>02-Feb-08 to 18-Mar-08 (45 days)</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>WR3</td>
<td>01-Mar-08</td>
<td>02-Apr-08 to 17-May-08 (45 days)</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>WR4</td>
<td>01-Mar-08</td>
<td>01-Jun-08 to 16-Jul-08 (45 days)</td>
<td>Yes</td>
<td>01-Jun-08 to 16-Jul-08</td>
</tr>
<tr>
<td>WR5</td>
<td>01-Mar-08</td>
<td>31-Jul-08 to 14-Sep-08 (45 days)</td>
<td>Yes</td>
<td>31-Jul-08 to 14-Sep-08 (45 days)</td>
</tr>
<tr>
<td>WR6</td>
<td>01-Mar-08</td>
<td>25-Nov-08 to 09-Jan-09 (45 days)</td>
<td>Yes</td>
<td>25-Nov-08 to 31-Dec-08</td>
</tr>
<tr>
<td>WR7</td>
<td>01-Mar-08</td>
<td>24-Jan-09 to 10-Mar-09 (45 days)</td>
<td>No</td>
<td>-</td>
</tr>
</tbody>
</table>

Note that in test case WR6, the lapse duration is 45 days; however, the expected lapse duration from the framework is 36 days. As discussed in earlier chapters, this is because for lapses that occur towards the end of the evaluation period, only that portion that overlaps with the evaluation period is reported.

The all-pairs test cases corresponding to the lapses in Figure 6.5 are shown in Table 6.2. In order to create the non-overlapping, nominal 45-day lapse durations within the six-month evaluation period, one needs to create prescriptions around these lapses, but these prescriptions may not necessarily have realistic durations. For example, to create the lapse pair for the test case AP16, three prescriptions need to be created: one from 01-Mar-08 to 01-Jun-08 (92 days), one from 16-Jul-08 to 31-Jul-08 (15 days) and another from 14-Sep-08 to 09-Jan-09 (117 days). However, the intention here is to create the two required lapse durations and as such this was considered an acceptable solution for testing purposes.
Table 6.2: All-pairs test cases for medication lapses with diagnosis prior to evaluation period

<table>
<thead>
<tr>
<th>Case ID</th>
<th>First Lapse Region</th>
<th>Second Lapse Region</th>
<th>Valid Lapse Region(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP1</td>
<td>1</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>AP 2</td>
<td>1</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>AP 3</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>AP 4</td>
<td>1</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>AP 5</td>
<td>1</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>AP 6</td>
<td>1</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>AP 7</td>
<td>2</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>AP 8</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>AP 9</td>
<td>2</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>AP 10</td>
<td>2</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>AP 11</td>
<td>2</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>AP 12</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>AP 13</td>
<td>3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>AP 14</td>
<td>3</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>AP 15</td>
<td>3</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>AP 16</td>
<td>4</td>
<td>5</td>
<td>4, 5</td>
</tr>
<tr>
<td>AP 17</td>
<td>4</td>
<td>6</td>
<td>4, 6</td>
</tr>
<tr>
<td>AP 18</td>
<td>4</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>AP 19</td>
<td>5</td>
<td>6</td>
<td>5, 6</td>
</tr>
<tr>
<td>AP 20</td>
<td>5</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>AP 21</td>
<td>6</td>
<td>7</td>
<td>6</td>
</tr>
</tbody>
</table>

There are 21 all-pairs test cases in Table 6.2. This is indeed the same number of test cases that results from equation 6.1, since there are seven distinct lapse locations (i.e., $n = 7$ in equation 6.1).

So far I have discussed only the cases where the diagnosis date was prior to the evaluation period. Another set of test cases is warranted when the diagnosis date is during the evaluation period, and these cases are shown in Figure 6.6.
The corresponding test cases for the scenarios in Figure 6.6 are shown in Table 6.3. Note that in WR11, the lapse duration is 40 days, but the expected output for this lapse is 35 days since this is the duration that falls within the evaluation period after the diagnosis date. Similar to WR6 in Figure 6.5, WR13 has a lapse duration of 45 days, but the expected output lapse duration is only 34 days. One may argue that cases 12, 13 and 14 in Figure 6.6 are redundant since these cases have the same relationships to the diagnosis date and the evaluation period, however, they have been included since they are important for all-pairs testing. For example, the two cases in AP17 in Table 6.2 are analogous to scenarios 9 and 13 in Figure 6.6 since they bear the same relationship to the evaluation period, but AP17 results in two valid lapses; in Figure 6.6, there is only one valid lapse. As such, cases 12, 13 and 14 are redundant in terms of weak-robust testing, but important for all-pairs testing (Table 6.4).
Table 6.3: Test cases for weak-robust equivalence class testing for medication lapse with diagnosis during the evaluation period

<table>
<thead>
<tr>
<th>Case ID</th>
<th>Diagnosis Date</th>
<th>Lapse Period</th>
<th>Is Valid Lapse?</th>
<th>Valid Lapse Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>WR8</td>
<td>01-Sep-08</td>
<td>30-Mar-08 to 14-May-08</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(45 days)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WR9</td>
<td>01-Sep-08</td>
<td>19-May-08 to 03-Jul-08</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(45 days)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WR10</td>
<td>01-Sep-08</td>
<td>08-Jul-08 to 22-Aug-08</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(45 days)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WR11</td>
<td>01-Sep-08</td>
<td>27-Aug-08 to 06-Oct-08</td>
<td>Yes</td>
<td>01-Sep-08 to 06-Oct-08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(40 days)</td>
<td></td>
<td>(35 days)</td>
</tr>
<tr>
<td>WR12</td>
<td>01-Sep-08</td>
<td>11-Oct-08 to 25-Nov-08</td>
<td>Yes</td>
<td>11-Oct-08 to 25-Nov-08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(45 days)</td>
<td></td>
<td>(45 days)</td>
</tr>
<tr>
<td>WR13</td>
<td>01-Sep-08</td>
<td>27-Nov-08 to 11-Jan-09</td>
<td>Yes</td>
<td>27-Nov-08 to 31-Dec-08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(45 days)</td>
<td></td>
<td>(34 days)</td>
</tr>
<tr>
<td>WR14</td>
<td>01-Sep-08</td>
<td>16-Jan-09 to 02-Mar-09</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(45 days)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.4 shows the test cases for all-pairs testing when the diagnosis date is during the evaluation period.
### Table 6.4: All-pairs test cases for medication lapses with diagnosis during the evaluation period

<table>
<thead>
<tr>
<th>Case ID</th>
<th>First Lapse Region</th>
<th>Second Lapse Region</th>
<th>Valid Lapse Region(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR1</td>
<td>8</td>
<td>9</td>
<td>-</td>
</tr>
<tr>
<td>SR2</td>
<td>10</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>SR3</td>
<td>11</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>SR4</td>
<td>12</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>SR5</td>
<td>13</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>SR6</td>
<td>14</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>SR7</td>
<td>9</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>SR8</td>
<td>11</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>SR9</td>
<td>12</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>SR10</td>
<td>13</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>SR11</td>
<td>14</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>SR12</td>
<td>10</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>SR13</td>
<td>12</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>SR14</td>
<td>13</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>SR15</td>
<td>14</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>SR16</td>
<td>11</td>
<td>12</td>
<td>11, 12</td>
</tr>
<tr>
<td>SR17</td>
<td>13</td>
<td>11, 13</td>
<td></td>
</tr>
<tr>
<td>SR18</td>
<td>14</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>SR19</td>
<td>12</td>
<td>13</td>
<td>12, 13</td>
</tr>
<tr>
<td>SR20</td>
<td>14</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>SR21</td>
<td>13</td>
<td>14</td>
<td>13</td>
</tr>
</tbody>
</table>

Note that Table 6.2 and Table 6.4 each has 21 test cases since all-pairs testing was used instead of exhaustive testing where all possible combinations need to be considered. If exhaustive testing was performed, this would require $2^n - 1$ (=127) test cases – over six times the number of test cases developed. The empty set in this calculation has been excluded since an empty set is not a valid test case in the context of all-pairs testing as it would mean ‘a no lapse at all’ condition which cannot be combined with any other case. In fact, this case is discussed as a special test case later.

When testing for medication lapses, we need to consider a few special cases not covered by the systematic identification of the equivalence classes presented so far, but rather based on domain expertise. This is because there are patients who have no medication lapse at all, or have a lapse that spans the duration with different relationships to the ones discussed so far. These special scenarios are depicted in Figure 6.7 and Figure 6.8.
Figure 6.7: Special cases for medication lapse when diagnosis is prior to the EP

Figure 6.8: Special cases for medication lapse when diagnosis is during the EP

The test cases corresponding to these scenarios are shown in Table 6.5. Note that for all the valid lapses except for SP16, the expected duration is from the start of the evaluation period or the diagnosis date, whichever is the latest, to the end of the evaluation period.
It is important to test that the statement – “expected duration of the valid lapse is from the start of the evaluation period or the diagnosis date, whichever is the latest” holds true even around the boundaries of the evaluation period. Since we need to determine lapses that are at least 30 days long, testing for this condition with the diagnosis date around the end of the evaluation period is not logically meaningful since the lapse will finish after the end of the evaluation period (WR6 and WR7 for example). Also, we have already tested the conditions where the diagnosis date was both well before and well after the beginning of the evaluation period (scenarios in Figure 6.5 and Figure 6.6 respectively). Therefore, the remaining test cases are when the diagnosis date is just before, equal to and just after the beginning of the evaluation period. The test cases for these scenarios are shown in Table 6.6.

Test case BV3 shows how the expected lapse duration starts on the diagnosis date while for the other two cases they start on the beginning of the evaluation period. Based on these test cases, it
should be clear that whenever a valid lapse occurs, the duration of the lapse starts from the start of the evaluation period or the diagnosis date, whichever is the latest.

Another interesting aspect that warrants consideration is how the interaction between a lapse period and the diagnosis date or the start of the evaluation period behaves around the boundaries. Scenarios 23-25 in Figure 6.9 depict cases where the lapse duration finishes just before, on and just after the start of the evaluation period, and scenarios 26-28 in Figure 6.10 show instances where the lapse period starts just after, on and just before the diagnosis date. All lapse durations are greater than or equal to the nominal 45 days so that they satisfy the minimum 30-day lapse duration requirement of the criterion. BVA around the diagnosis date when it is prior to the beginning of the evaluation period is of little or no special interest when considering nominal lapse durations that overlap with the evaluation period since all of them will be valid scenarios, hence only the cases when diagnosis is during the evaluation period are tested (Figure 6.10).

![Figure 6.9: BVA on the interaction between the lapse period and the beginning of the evaluation period](image1)

![Figure 6.10: BVA on the interaction between the lapse period and the diagnosis date](image2)
Table 6.7 shows the test cases for the scenarios in Figure 6.9 and Figure 6.10 with the input parameters and expected output details. It is important to realise that BV24 is not deemed to be a valid lapse duration because this lapse finishes on the start date of the evaluation period and therefore does not run-into the evaluation period. Also, note the difference in the durations between the input lapse and the expected output lapse for BV28.

Table 6.7: BVA test cases for the interaction between the lapse period and the diagnosis date or beginning of the evaluation period

<table>
<thead>
<tr>
<th>Case ID</th>
<th>Diagnosis Date</th>
<th>Lapse Period</th>
<th>Is Valid Lapse?</th>
<th>Valid Lapse Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>BV23</td>
<td>01-May-08</td>
<td>16-May-08 to 30-Jun-08 (45 days)</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>BV24</td>
<td>01-May-08</td>
<td>17-May-08 to 01-Jul-08 (45 days)</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>BV25</td>
<td>01-May-08</td>
<td>18-May-08 to 02-Jul-08 (45 days)</td>
<td>Yes</td>
<td>18-May-08 to 02-Jul-08 (45 days)</td>
</tr>
<tr>
<td>BV26</td>
<td>01-Sep-08</td>
<td>02-Sep-08 to 17-Oct-08 (45 days)</td>
<td>Yes</td>
<td>02-Sep-08 to 17-Oct-08 (45 days)</td>
</tr>
<tr>
<td>BV27</td>
<td>01-Sep-08</td>
<td>01-Sep-08 to 16-Oct-08 (45 days)</td>
<td>Yes</td>
<td>01-Sep-08 to 16-Oct-08 (45 days)</td>
</tr>
<tr>
<td>BV28</td>
<td>01-Sep-08</td>
<td>31-Aug-08 to 15-Oct-08 (45 days)</td>
<td>Yes</td>
<td>01-Sep-08 to 16-Oct-08 (44 days)</td>
</tr>
</tbody>
</table>

The various test cases related to equivalence class testing have been discussed so far, namely, single fault testing, all-pairs testing and some BVA test cases as applicable to the relationship between the diagnosis date and the beginning of the evaluation period (i.e., the location of the lapse). It is equally important to perform BVA for the duration of the lapse, and Figure 6.11 shows the different test cases that need to be considered for this analysis. The diagnosis date in this case is assumed to be prior to the start of the evaluation period.
It is interesting to note the differences and the similarities between Figure 6.11 and Figure 6.4 (the latter was a function $F(x_1, x_2)$ with $x_1$ having the range $[a, b]$ and $x_2$ having the range $[c, d]$). In Figure 6.11, there is a single threshold value as opposed to the range $[c, d]$ and three descriptive boundaries are used instead of a numerical boundary range $[a, b]$. However, the relevance of Figure 6.11 to robustness testing should be clear. The time-transformed graphical representation of the scenarios in Figure 6.11 is shown in Figure 6.12.
The input parameters and the expected output lapse details for the test cases for BVA on lapse duration are shown in Table 6.8.

<table>
<thead>
<tr>
<th>Case ID</th>
<th>Diagnosis Date</th>
<th>Lapse Period</th>
<th>Is Valid Lapse?</th>
<th>Valid Lapse Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>BV29</td>
<td>01-Jun-08</td>
<td>25-Jun-08 to 10-Jul-08 (15 days)</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>BV30</td>
<td>01-Jun-08</td>
<td>25-Jun-08 to 24-Jul-08 (29 days)</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>BV31</td>
<td>01-Jun-08</td>
<td>25-Jun-08 to 25-Jul-08 (30 days)</td>
<td>Yes</td>
<td>25-Jun-08 to 25-Jul-08 (30 days)</td>
</tr>
<tr>
<td>BV32</td>
<td>01-Jun-08</td>
<td>25-Jun-08 to 26-Jul-08 (31 days)</td>
<td>Yes</td>
<td>25-Jun-08 to 26-Jul-08 (31 days)</td>
</tr>
<tr>
<td>BV33</td>
<td>01-Jun-08</td>
<td>25-Jun-08 to 09-Aug-08 (45 days)</td>
<td>Yes</td>
<td>25-Jun-08 to 09-Aug-08 (45 days)</td>
</tr>
<tr>
<td>BV34</td>
<td>01-Jun-08</td>
<td>01-Sep-08 to 16-Sep-08 (15 days)</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>BV35</td>
<td>01-Jun-08</td>
<td>01-Sep-08 to 30-Sep-08 (29 days)</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>BV36</td>
<td>01-Jun-08</td>
<td>01-Sep-08 to 01-Oct-08 (30 days)</td>
<td>Yes</td>
<td>01-Sep-08 to 01-Oct-08 (30 days)</td>
</tr>
<tr>
<td>BV37</td>
<td>01-Jun-08</td>
<td>01-Sep-08 to 02-Oct-08 (31 days)</td>
<td>Yes</td>
<td>01-Sep-08 to 02-Oct-08 (31 days)</td>
</tr>
<tr>
<td>BV38</td>
<td>01-Jun-08</td>
<td>01-Sep-08 to 16-Oct-08 (45 days)</td>
<td>Yes</td>
<td>01-Sep-08 to 16-Oct-08 (45 days)</td>
</tr>
<tr>
<td>BV39</td>
<td>01-Jun-08</td>
<td>16-Dec-08 to 14-Feb-09 (60 days)</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>BV40</td>
<td>01-Jun-08</td>
<td>02-Dec-08 to 31-Jan-09 (60 days)</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>BV41</td>
<td>01-Jun-08</td>
<td>01-Dec-08 to 30-Jan-09 (60 days)</td>
<td>Yes</td>
<td>01-Dec-08 to 31-Dec-08 (30 days)</td>
</tr>
<tr>
<td>BV42</td>
<td>01-Jun-08</td>
<td>30-Nov-08 to 29-Jan-09 (60 days)</td>
<td>Yes</td>
<td>30-Nov-08 to 31-Dec-08 (31 days)</td>
</tr>
<tr>
<td>BV43</td>
<td>01-Sep-08</td>
<td>16-Nov-08 to 15-Jan-09 (60 days)</td>
<td>Yes</td>
<td>16-Nov-08 to 31-Dec-08 (45 days)</td>
</tr>
</tbody>
</table>

So far, when analysing a test case for a lapse period, the period was referred to in terms of a medication lapse, although the test cases required to test the timely measurement recording criterion are essentially the same. A measurement lapse can occur in any of the ways discussed in this section, and once a measurement lapse has been identified, the software module that processes lapse periods with respect to a reference date (diagnosis date or a high BP measurement for instance) and a reference period (i.e., the evaluation period) processes the lapse in the same manner, irrespective of whether it was for a medication lapse or a measurement lapse. As such, there is no need to discuss the same set of test cases again for measurement lapses.
6.3.2 Test Cases for Measurement Contraindicating Therapy

The key temporal requirement to satisfying this criterion is having a measurement of the required type within the evaluation period while being on a specific therapy (often after a particular diagnosis). The equivalence classes identified for this criterion are shown in Figure 6.13. Here we assume that the threshold value of the lab result satisfies the criterion requirement (eGFR > 60 mL/min for example). Also, the software module that determines whether a period (whether it is a lapse period or a period of medication coverage) has an overlap with the evaluation period is the same as the one used to test prior criteria, and therefore has already been thoroughly tested previously (per Figure 6.5, Figure 6.6 and Figure 6.9 for example). Therefore, much of the testing required in this section has already been tested and is not duplicated here. For example, Figure 6.13 does not test whether the medication coverage was prior to the beginning of the evaluation period.

![Diagram](image)

**Figure 6.13:** Weak-robust equivalence class testing for measurement contraindicating therapy. The indicated durations represent medication coverage.

The test cases corresponding to the weak-robust equivalence classes in Figure 6.13 are shown in Table 6.9. Only test case WR44 satisfies all the criteria required to satisfy this audit criterion. WR45 and WR46 both have the medication coverage and the measurement date during the evaluation period, but these cases fail as the measurement date does not occur during the period of medication coverage. The medication coverages of WR47 and WR48 contain the lab measurement date, however they fail to satisfy the audit criterion since the measurement date falls before the start of the evaluation period and after the end of the evaluation period respectively.
### Table 6.9: Weak-robust equivalence class test cases for the management of compelling indications

<table>
<thead>
<tr>
<th>Case ID</th>
<th>Diagnosis Date</th>
<th>Medication Coverage</th>
<th>Measurement Date</th>
<th>Satisfies Criterion?</th>
</tr>
</thead>
<tbody>
<tr>
<td>WR44</td>
<td>01-May-08</td>
<td>01-Sep-08 to 16-Oct-08 (45 days)</td>
<td>01-Oct-08</td>
<td>Yes</td>
</tr>
<tr>
<td>WR45</td>
<td>01-May-08</td>
<td>01-Sep-08 to 16-Oct-08 (45 days)</td>
<td>01-Aug-08</td>
<td>No</td>
</tr>
<tr>
<td>WR46</td>
<td>01-May-08</td>
<td>01-Sep-08 to 16-Oct-08 (45 days)</td>
<td>01-Nov-08</td>
<td>No</td>
</tr>
<tr>
<td>WR47</td>
<td>01-May-08</td>
<td>25-Jun-08 to 09-Aug-08 (45 days)</td>
<td>28-Jun-08</td>
<td>No</td>
</tr>
<tr>
<td>WR48</td>
<td>01-May-08</td>
<td>16-Nov-08 to 15-Jan-09 (60 days)</td>
<td>10-Jan-09</td>
<td>No</td>
</tr>
</tbody>
</table>

Figure 6.14 and Figure 6.15 show the scenarios for BVA around the start and end of the evaluation period as well as the diagnosis date. Performing BVA around the diagnosis date in Figure 6.14 will be redundant because the lab measurement date would be before the start of the evaluation period; hence a separate figure is used.

![Figure 6.14: BVA on lab measurement date with respect to the start and the end of the evaluation period](image)

Figure 6.14: BVA on lab measurement date with respect to the start and the end of the evaluation period
Table 6.10 shows the test cases with input and expected output details for the scenarios shown in Figure 6.14 and Figure 6.15.

<table>
<thead>
<tr>
<th>Case ID</th>
<th>Diagnosis Date</th>
<th>Medication Coverage</th>
<th>Measurement Date</th>
<th>Satisfies Criterion?</th>
</tr>
</thead>
<tbody>
<tr>
<td>WR49</td>
<td>01-May-08</td>
<td>25-Jun-08 to 09-Aug-08 (45 days)</td>
<td>30-Jun-08</td>
<td>No</td>
</tr>
<tr>
<td>WR50</td>
<td>01-May-08</td>
<td>25-Jun-08 to 09-Aug-08 (45 days)</td>
<td>01-Jul-08</td>
<td>Yes</td>
</tr>
<tr>
<td>WR51</td>
<td>01-May-08</td>
<td>25-Jun-08 to 09-Aug-08 (45 days)</td>
<td>02-Jul-08</td>
<td>Yes</td>
</tr>
<tr>
<td>WR52</td>
<td>01-May-08</td>
<td>30-Nov-08 to 29-Jan-09 (60 days)</td>
<td>30-Dec-08</td>
<td>Yes</td>
</tr>
<tr>
<td>WR53</td>
<td>01-May-08</td>
<td>30-Nov-08 to 29-Jan-09 (60 days)</td>
<td>31-Dec-08</td>
<td>Yes</td>
</tr>
<tr>
<td>WR54</td>
<td>01-May-08</td>
<td>30-Nov-08 to 29-Jan-09 (60 days)</td>
<td>01-Jan-08</td>
<td>No</td>
</tr>
<tr>
<td>WR55</td>
<td>01-Sep-08</td>
<td>01-Aug-08 to 15-Sep-08 (45 days)</td>
<td>31-Aug-08</td>
<td>No</td>
</tr>
<tr>
<td>WR56</td>
<td>01-Sep-08</td>
<td>01-Aug-08 to 15-Sep-08 (45 days)</td>
<td>01-Sep-08</td>
<td>Yes</td>
</tr>
<tr>
<td>WR57</td>
<td>01-Sep-08</td>
<td>01-Aug-08 to 15-Sep-08 (45 days)</td>
<td>02-Sep-08</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Equivalence class testing and BVA have been carried out so far, but based on domain expertise, there are also a few special cases that warrant consideration. These cases are shown in Figure 6.16 and Figure 6.17, and the corresponding test case details are shown in Table 6.10.
Figure 6.16: Special value testing for measurement contraindicating therapy

Figure 6.17: Special value testing for measurement contraindicating therapy with diagnosis date during the evaluation period
<table>
<thead>
<tr>
<th>Case ID</th>
<th>Diagnosis Date</th>
<th>Medication Coverage</th>
<th>Measurement Date</th>
<th>Satisfies Criterion?</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP58</td>
<td>01-May-08</td>
<td>-</td>
<td>-</td>
<td>No</td>
</tr>
<tr>
<td>SP59</td>
<td>01-May-08</td>
<td>01-Sep-08 to 16-Oct-08 (45 days)</td>
<td>-</td>
<td>No</td>
</tr>
<tr>
<td>SP60</td>
<td>01-May-08</td>
<td>01-Oct-08</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>SP61</td>
<td>01-May-08</td>
<td>01-Jun-08 to 01-Jul-08 (30 days)</td>
<td>01-Jul-08</td>
<td>No</td>
</tr>
<tr>
<td>SP62</td>
<td>01-May-08</td>
<td>01-Jul-08 to 01-Aug-08 (31 days)</td>
<td>01-Jul-08</td>
<td>Yes</td>
</tr>
<tr>
<td>SP63</td>
<td>01-May-08</td>
<td>01-Dec-08 to 31-Dec-08 (30 days)</td>
<td>31-Dec-08</td>
<td>Yes</td>
</tr>
<tr>
<td>SP64</td>
<td>01-May-08</td>
<td>31-Dec-08 to 30-Jan-09 (30 days)</td>
<td>31-Dec-08</td>
<td>No</td>
</tr>
<tr>
<td>SP65</td>
<td>01-Sep-08</td>
<td>02-Aug-08 to 01-Sep-08 (30 days)</td>
<td>01-Sep-08</td>
<td>Yes</td>
</tr>
<tr>
<td>SP66</td>
<td>01-Sep-08</td>
<td>01-Sep-08 to 01-Oct-09 (30 days)</td>
<td>01-Sep-08</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Figure 6.18: Test cases for measurement contraindicating therapy

SP58 is a case where there is neither medication coverage nor a valid lab test. SP59 has a valid medication coverage, but no lab measurement, while SP60 has the lab measurement, but no medication coverage. Testing cases 61-66 is somewhat similar to performing BVA on multiple variables since we are testing the interaction of the lab measurement date and the start or end of the medication coverage with the boundaries of the evaluation period and the diagnosis date. However, these cases have been developed based on domain knowledge rather than formal BVA on multiple variables, and therefore are considered here under special value testing.

### 6.3.3 Test Cases for Time to Achieve Target

This criterion requires a specified number of uncontrolled measurements to occur over a given minimum period of time either with the last uncontrolled measurement occurring during the evaluation period or else with no subsequently controlled measurement occurring after the uncontrolled measurements. The weak-robust equivalence classes for this criterion are shown in Figure 6.19. Note that for testing purposes, BP was used as the measurement type of interest, and 160/100 mmHg was defined as an ‘uncontrolled’ measurement. At least three consecutively high BPs over at least 120 days were required to satisfy the criterion.
Scenario 67 in Figure 6.19 represents a typical case where the last uncontrolled measurement was during the evaluation period. This case will be valid irrespective of whether there was controlled measurement during the evaluation period or not. Scenario 68 represents a case where the consecutively high BPs occurred prior to the evaluation period, with no controlled BP measurement occurring afterward; and therefore needs to be identified as a valid case. Scenario 69 is similar to scenario 68, but there is a subsequently controlled BP measurement appears after the uncontrolled measurements, and hence Scenario 69 is not a valid case. Note that whether the controlled measurement occurs before or after the start of the evaluation period is not important, and the only requirement is for this to be after the end of the last uncontrolled measurement and before the end of the evaluation period. Since neither scenario 70 nor 71 satisfies the minimum number of uncontrolled measurements that need to be present during the evaluation period, they are not valid cases. These test cases are tabulated in Table 6.11 with input parameter and expected output details.
<table>
<thead>
<tr>
<th>Case ID</th>
<th>Diagnosis Date</th>
<th>Measurement Number in Sequence</th>
<th>Measurement Date</th>
<th>Satisfies Criterion?</th>
</tr>
</thead>
<tbody>
<tr>
<td>WR67</td>
<td>01-Feb-08</td>
<td>WR67_1</td>
<td>01-Mar-08</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR67_2</td>
<td>01-May-08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR67_3</td>
<td>10-Jul-08</td>
<td></td>
</tr>
<tr>
<td>WR68</td>
<td>01-Feb-08</td>
<td>WR68_1</td>
<td>15-Feb-08</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR68_2</td>
<td>01-May-08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR68_3</td>
<td>25-Jun-08</td>
<td></td>
</tr>
<tr>
<td>WR69</td>
<td>01-Feb-08</td>
<td>WR69_1</td>
<td>15-Feb-08</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR69_2</td>
<td>01-May-08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR69_3</td>
<td>25-Jun-08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR69_4 (controlled)</td>
<td>25-Jul-08</td>
<td></td>
</tr>
<tr>
<td>WR70</td>
<td>01-Feb-08</td>
<td>WR70_1</td>
<td>20-Aug-08</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR70_2</td>
<td>01-Nov-08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR70_3</td>
<td>15-Jan-09</td>
<td></td>
</tr>
<tr>
<td>WR71</td>
<td>01-Feb-08</td>
<td>WR71_1</td>
<td>10-Jan-09</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR71_2</td>
<td>01-Mar-09</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR71_3</td>
<td>01-Jun-09</td>
<td></td>
</tr>
</tbody>
</table>

Other than equivalence class testing, it is important to perform BVA to ensure that the measurements around the important boundaries such as the start and end date of the evaluation period are handled correctly. Table 6.12 shows the BVA test cases around the start and end of the evaluation period as well as the diagnosis date when it is during the evaluation period. Note that expected outcomes from scenarios BV78, BV79 and BV80 would have been similar even if the diagnosis was prior to the evaluation period, and there was no subsequently controlled measurement. All test cases have a difference greater than 120 days between the first measurement and the last measurement within a sequence.
<table>
<thead>
<tr>
<th>Case ID</th>
<th>Diagnosis Date</th>
<th>Measurement Number in Sequence</th>
<th>Measurement Date</th>
<th>Expected Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>BV72</td>
<td>01-Feb-08</td>
<td>BV72_1</td>
<td>01-Mar-08</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BV72_2</td>
<td>01-May-08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BV72_3</td>
<td>30-Jun-08</td>
<td></td>
</tr>
<tr>
<td>BV73</td>
<td>01-Feb-08</td>
<td>BV73_1</td>
<td>01-Mar-08</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BV73_2</td>
<td>01-May-08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BV73_3</td>
<td>01-Jul-08</td>
<td></td>
</tr>
<tr>
<td>BV74</td>
<td>01-Feb-08</td>
<td>BV74_1</td>
<td>01-Mar-08</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BV74_2</td>
<td>01-May-08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BV74_3</td>
<td>02-Jul-08</td>
<td></td>
</tr>
<tr>
<td>BV75</td>
<td>01-Feb-08</td>
<td>WR75_1</td>
<td>20-Aug-08</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR75_2</td>
<td>01-Nov-08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR75_3</td>
<td>30-Dec-08</td>
<td></td>
</tr>
<tr>
<td>BV76</td>
<td>01-Feb-08</td>
<td>WR76_1</td>
<td>20-Aug-08</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR76_2</td>
<td>01-Nov-08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR76_3</td>
<td>31-Dec-08</td>
<td></td>
</tr>
<tr>
<td>BV77</td>
<td>01-Feb-08</td>
<td>WR77_1</td>
<td>20-Aug-08</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR77_2</td>
<td>01-Nov-08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR77_3</td>
<td>01-Jan-09</td>
<td></td>
</tr>
<tr>
<td>BV78</td>
<td>01-Aug-08</td>
<td>WR78_1</td>
<td>31-Jul-08</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR78_2</td>
<td>01-Oct-08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR78_3</td>
<td>12-Dec-08</td>
<td></td>
</tr>
<tr>
<td>BV79</td>
<td>01-Aug-08</td>
<td>WR79_1</td>
<td>01-Aug-08</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR79_2</td>
<td>01-Oct-08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR79_3</td>
<td>12-Dec-08</td>
<td></td>
</tr>
<tr>
<td>BV80</td>
<td>01-Aug-08</td>
<td>WR80_1</td>
<td>02-Aug-08</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR80_2</td>
<td>01-Oct-08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WR80_3</td>
<td>12-Dec-08</td>
<td></td>
</tr>
</tbody>
</table>

Two other aspects of BVA that need to be performed are (i) the behaviour around the minimum required number of days over which the consecutively uncontrolled measurements should occur and (ii) the minimum number of measurements. The test cases corresponding to these two are shown in Table 6.13 and Table 6.14 respectively.
Table 6.13: BVA around the minimum required number of days over which consecutively uncontrolled measurements should occur

<table>
<thead>
<tr>
<th>Case ID</th>
<th>Diagnosis Date</th>
<th>Measurement Number in Sequence</th>
<th>Measurement Date</th>
<th>Difference between First and Last Measurement in Sequence in Days</th>
<th>Satisfies Criterion?</th>
</tr>
</thead>
<tbody>
<tr>
<td>BV81</td>
<td>01-Feb-08</td>
<td>BV81_1</td>
<td>10-Jul-08</td>
<td>119</td>
<td>No</td>
</tr>
<tr>
<td>BV81</td>
<td></td>
<td>BV81_2</td>
<td>01-Sep-08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BV81</td>
<td></td>
<td>BV81_3</td>
<td>06-Nov-08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BV82</td>
<td>01-Feb-08</td>
<td>BV82_1</td>
<td>10-Jul-08</td>
<td>120</td>
<td>Yes</td>
</tr>
<tr>
<td>BV82</td>
<td></td>
<td>BV82_2</td>
<td>01-Sep-08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BV82</td>
<td></td>
<td>BV82_3</td>
<td>07-Nov-08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BV83</td>
<td>01-Feb-08</td>
<td>BV83_1</td>
<td>10-Jul-08</td>
<td>121</td>
<td>Yes</td>
</tr>
<tr>
<td>BV83</td>
<td></td>
<td>BV83_2</td>
<td>01-Sep-08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BV83</td>
<td></td>
<td>BV83_3</td>
<td>08-Nov-08</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.14: BVA around the minimum number of measurements required to satisfy criterion

<table>
<thead>
<tr>
<th>Case ID</th>
<th>Diagnosis Date</th>
<th>Measurement Number in Sequence</th>
<th>Measurement Date</th>
<th>Satisfies Criterion?</th>
</tr>
</thead>
<tbody>
<tr>
<td>BV84</td>
<td>01-Feb-08</td>
<td>BV84_1</td>
<td>10-Jul-08</td>
<td>No</td>
</tr>
<tr>
<td>BV84</td>
<td></td>
<td>BV84_2</td>
<td>28-Nov-08</td>
<td></td>
</tr>
<tr>
<td>BV85</td>
<td>01-Feb-08</td>
<td>BV85_1</td>
<td>10-Jul-08</td>
<td>Yes</td>
</tr>
<tr>
<td>BV85</td>
<td></td>
<td>BV85_2</td>
<td>01-Sep-08</td>
<td></td>
</tr>
<tr>
<td>BV85</td>
<td></td>
<td>BV85_3</td>
<td>20-Oct-08</td>
<td></td>
</tr>
<tr>
<td>BV85</td>
<td></td>
<td>BV85_4</td>
<td>28-Nov-08</td>
<td></td>
</tr>
</tbody>
</table>

BV81 represents a case where the consecutively uncontrolled measurements occur over 119 days, hence is not flagged as a valid case. BV82 and BV83 both satisfy all of the required criteria, and are therefore valid cases.

BV84 is a test case with only two consecutively uncontrolled measurements, and is not a valid case. On the other hand, BV85 has four consecutively uncontrolled measurements, and therefore satisfies the audit criterion. A test case with exactly three measurements is not used here as all the previous test cases for this criterion consisted of three measurements. Note that from a strict BVA testing point of view, all the previous test cases should be tested with a nominal number of consecutively uncontrolled measurements, such as five measurements. The boundary number of test cases, that is, two, three and four should be tested only when performing BVA around the minimum number of required measurements as shown in Table 6.14. However, graphically representing the different scenarios as in Figure 6.19 and representing the test case details as in Table 6.11 and Table 6.12 is clearer with only three...
consecutively uncontrolled measurements. Therefore, developing the test cases with three high BPs was deemed a suitable approach for this criterion class.

6.4 Random Testing to Ensure Software Reliability

Random testing is a form of functional testing that refers to selecting test inputs randomly from the input domain of a system [266]. As mentioned at the beginning of this chapter, testing any software system of reasonable complexity for all possible input combinations requires a great amount of effort. Even with only seven equivalence classes, in Section 6.3.1 all-pairs testing was used instead of exhaustive testing in order to reduce the number of test cases required to perform an acceptable level of testing. Therefore, the various test cases discussed so far represent only a subset of all possible test cases. Performing random testing with a large number of test cases has the potential to cover all possible combinations of the input domain (with some probability of each combination occurring), and has been suggested as a possible solution to ensure software reliability [247, 266].

When discussing random testing, Hamlet [247] states that:

“[i]n systematic testing, some aspect of a program is identified as a potential source of failure and the method attempts to make sure that this aspect is not neglected in testing. If a statement has never been executed during testing, for example, it could be radically wrong, yet its fault would not show. Or, a required software function could be implemented very badly – perhaps not implemented at all – yet the error will not be detected if testing fails to invoke that function. Random testing, on the other hand, makes no claims to cover anything, except insofar as chance dictates...There is an intuitive explanation for the surprising success of random testing relative to its systematic competitors. Systematic methods are more subjective than they appear, because in most cases there are an infinite number of ways to satisfy the method, and the choice is usually made by a human being. There are good and bad choices, relative to finding failures. Random testing can be viewed as replacing human choice with chance selection, including the possibility of no selection at all. For example, randomly selected test points may invoke all functions, all statements, etc., but the possibility always exists that some such elements could be missed”.
As such, random testing is not suggested as an alternative to systematic testing (which is usually performed to detect specific faults), but rather as a complementary testing scheme that has the potential to use the high number of ways the program’s input domain can be partitioned.

An issue with random testing however, is that it requires some kind of automation to generate the large number of test cases and also compute the expected outcomes [247, 266]. In an early survey on testing methods, Howden [264] states that “the use of testing requires the existence of an external mechanism which can be used to check the test output for correctness. This mechanism is referred to as the test oracle”, a term that has commonly appeared since in the testing literature. Several types of test oracles can be designed [263], but for the purposes of this work a specialisation of the gold standard oracle (gold standard oracles use one or more versions of an existing system to generate expected results) called a trusted system oracle has been used (Figure 6.20), which is usually an existing system that is used with high confidence.

![Figure 6.20: Components of a conceptual trusted system gold standard oracle [263]](image)

A test oracle acts (i) as a result generator for generating the expected results for the test inputs and (ii) as a result comparator for comparing the expected results with the actual results after executing a test case [266]. The two implementations developed for ChronoMedIt have been discussed previously (i.e., the SQL based implementation and the sequential based implementation) and the various testing performed so far has been on the sequential implementation. Therefore, for random testing purposes, the sequential implementation was considered to be the golden implementation that would produce the defect-free results that the SQL based implementation results could be compared against. The term test harness has been used to refer to a system that supports effective and repeatable automated testing, initiates the tests and then evaluates the test results [268]. This makes the test harness a central part of any test environment. The test harness in the context of this framework involves executing the two
independent implementations and then comparing the test results for a given criterion. This process overview is shown in Figure 6.21.

![Figure 6.21: Executing the SQL and sequential based implementations for test cases and comparing results for a given criterion](image)

A concept closely associated with generating random test cases is the operational profile of the input space [247, 248] which is simply a set of disjoint (only one can occur at a time) alternatives with the probability that each will occur [248]. The operational profile should account for usage patterns that the software will encounter in its intended environment. The argument here is that by testing according to usage, the failures found by imperfect methods are more likely to be the important ones, that is, the ones most users would encounter first [247]. It has been suggested that when defining this profile, a practically complete list of input variables from the input space should be chosen [248]. Taking this design consideration into account, the operational profile used for framework testing is shown in Table 6.15.
Table 6.15: Operational profile for various parameters in the input space

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value/Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of test patients</td>
<td>10,000</td>
</tr>
<tr>
<td>Number of prescriptions per patient</td>
<td>[0, 20]</td>
</tr>
<tr>
<td>Duration of a prescription</td>
<td>[15, 390]</td>
</tr>
<tr>
<td>Prescription issue dates</td>
<td>[01-July-08, 31-December-09]</td>
</tr>
<tr>
<td>Number of diagnoses per patient</td>
<td>[0, 5]</td>
</tr>
<tr>
<td>Diagnoses dates</td>
<td>[01-Jan-05, 31-December-09]</td>
</tr>
<tr>
<td>Number of BP measurements per patient</td>
<td>[0, 25]</td>
</tr>
<tr>
<td>Systolic BP value</td>
<td>[80, 200]</td>
</tr>
<tr>
<td>Diastolic BP value</td>
<td>[40, 140]</td>
</tr>
<tr>
<td>BP measurement dates</td>
<td>[01-July-08, 31-December-09]</td>
</tr>
<tr>
<td>Number of eGFR measurements per patient</td>
<td>[0, 8]</td>
</tr>
<tr>
<td>eGFR value</td>
<td>[0, 100]</td>
</tr>
<tr>
<td>eGFR measurement dates</td>
<td>[01-July-08, 31-December-09]</td>
</tr>
<tr>
<td>Number of uric acid measurements per patient</td>
<td>[0, 8]</td>
</tr>
<tr>
<td>Uric acid value</td>
<td>[0, 6]</td>
</tr>
<tr>
<td>Uric acid measurement dates</td>
<td>[01-July-08, 31-December-09]</td>
</tr>
</tbody>
</table>

Values for all parameters except prescription duration were generated using a discrete uniform distribution. For a given range \([a, b]\), the discrete uniform distribution is given by:

\[
(\_k\_g_{\_1}) = \begin{cases} 
\frac{1}{n}, & \text{if } k \neq 0 \\
0, \quad & \text{if } h
\end{cases}
\]

where \(k\) denotes a discrete number out of the \(n\) possible values.

Prescriptions, however, are usually issued as 90-day scripts, so using a uniform distribution will not result in realistic prescription durations. To overcome this situation, a beta distribution was used to determine the prescription durations with a mode of 90 days. The beta distribution is a continuous probability distribution defined on the interval \([0, 1]\) and parameterised by two positive shape parameters, typically denoted by \(\alpha\) and \(\beta\). In order to obtain the shape required to represent prescription durations, these two parameters were set to 2 and 5 respectively. The random numbers following this distribution on the interval \([0, 1]\) were generated using the \textit{NMath Stats} library (which requires the \textit{NMath Core} library) available as a .NET based component from CentreSpace [193].

\[2\] It was ensured that the diastolic BP was lower than the corresponding systolic BP.
The generalised beta distribution has the range \([a, b]\) and has extra parameters that define the minimum and maximum. The mode \(m\) is then given by:

\[
  m = a + (b - a) \frac{-1}{\alpha + \beta - 2}
\]  

(6.3)

The minimum duration of a script was set to be 15 days based on experience with analysing production EMR prescribing data. When \(m, a, \alpha\) and \(\beta\) are set in Eq. 6.3, \(b\) results in 390 days which was then used as the maximum duration for a given script. This is a reasonable maximum duration for a script when the duration of the evaluation period is less than 390 days (which is often the case) and it denotes a case similar to SP17, SP18 and/or SP21 in Table 6.5.

The random numbers generated on the interval \([0, 1]\) were then transformed to \([15, 390]\) using a linear transformation of \([(390-15)x + 15]\) and then rounded to the nearest day, which results in the distribution shown in Figure 6.22.

![Figure 6.22: Operational profile for prescription durations](image)

Table 6.15 indicated that the number of prescriptions per patient was \([0, 20]\), and for a large number of patients such as 10,000, the total number of prescriptions averages out to be approximately 100,000. In fact, Figure 6.22 was generated using 99,458 such data points.
Other than the duration, each prescription also has an associated generic drug, and this was randomly selected from the pool of available generic drugs. For example, if the drug knowledge base (Section 5.5.2) consisted of 125 generic drug names, a random number in the range [0, 124] was generated and the corresponding drug name was selected from the list (note that in C#, arrays are 0-indexed, and therefore 0 will correspond to the first generic drug). Similarly, a patient diagnosis requires a corresponding Read Clinical Code, and this was selected randomly from the pool of Read Clinical Codes from the classification knowledge base using a similar approach.

Details of the test database that was populated are shown in Table 6.16.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Number of Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>10,000</td>
</tr>
<tr>
<td>Number of prescriptions</td>
<td>99458</td>
</tr>
<tr>
<td>Number of diagnoses</td>
<td>25366</td>
</tr>
<tr>
<td>Number of BP measurements</td>
<td>126,315</td>
</tr>
<tr>
<td>Number of eGFR measurements</td>
<td>39748</td>
</tr>
<tr>
<td>Number of uric acid measurements</td>
<td>39361</td>
</tr>
</tbody>
</table>

Once the test database was populated with patient data, testing was carried out for the audit criteria shown in Table 6.17 for the one year evaluation period from 01-Jan-09 to 31-Dec-09 using a six-month run-in period. The resulting number of patients satisfying each criterion is also shown alongside.
Table 6.17: Audit criteria with parameters for random testing

<table>
<thead>
<tr>
<th>Audit Criterion Type</th>
<th>Key Parameters</th>
<th>Patients Satisfying Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persistence of medication</td>
<td>Medication: Antihypertensives Classification: Hypertension</td>
<td>219</td>
</tr>
<tr>
<td></td>
<td>Medications: ACEis or ARBs Classification: Hypertension</td>
<td>523</td>
</tr>
<tr>
<td></td>
<td>Medications: ACEis or ARBs Classification: Hypertension and diabetes</td>
<td>318</td>
</tr>
<tr>
<td></td>
<td>Medication: Cilazapril Classification: Diabetes</td>
<td>1457</td>
</tr>
<tr>
<td></td>
<td>Medication: Cilazapril Classification: Hypertension and diabetes</td>
<td>258</td>
</tr>
<tr>
<td></td>
<td>Medications: Cilazapril or quinapril Classification: Diabetes</td>
<td>1923</td>
</tr>
<tr>
<td></td>
<td>Medications: Cilazapril or quinapril Classification: Hypertension and diabetes</td>
<td>343</td>
</tr>
<tr>
<td>Timely measurement recording</td>
<td>Measurement: BP Classification: Hypertension</td>
<td>789</td>
</tr>
<tr>
<td></td>
<td>Minimum lapse duration: 120 days</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Measurement: BP Classification: Hypertension and diabetes Minimum lapse duration: 100 days After threshold: 140/90 mmHg</td>
<td>1033</td>
</tr>
<tr>
<td>Time to achieve target</td>
<td>Measurement: BP Classification: Hypertension Minimum consecutive measurements: 3 Threshold: 160/100 mmHg Measurements over days = 120</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Measurement: BP Classification: Hypertension and diabetes Minimum consecutive measurements: 4 Threshold: 140/90 mmHg Measurements over days = 150</td>
<td>102</td>
</tr>
<tr>
<td>Treatment of compelling indications</td>
<td>Measurement: eGFR Medication: Thiazide Classification: Hypertension and renal failure Value less than = 60</td>
<td>516</td>
</tr>
<tr>
<td></td>
<td>Measurement: Uric acid Medication: Beta blockers Classification: Hypertension and renal failure Value greater than = 0.42</td>
<td>503</td>
</tr>
</tbody>
</table>

Note that for persistence of medication criterion, the seven tests in Table 6.17 represent different combinations of single vs. multiple drug(s) or drug class(es) and classification(s). These tests were repeated for (i) lapse duration = 30, (ii) MPR ≥80%, (iii) MPR <80%, (iv) lapse duration = 30 and MPR ≥80%, and (v) lapse duration = 30 and MPR <80%, resulting in 35 different audit criteria tests. This was required since persistence of medication criterion can be specified in terms of lapse only, MPR only or as a combination of lapse and MPR. However, in Table 6.17, the number of patients corresponding only to ‘lapse duration = 30 and 0 < MPR <80%’ is shown...
as domain experience indicates that this is the most commonly used medication persistence criterion.

### 6.5 Errors Detected

Random testing for the cases shown in Table 6.17 was carried out following the iterative process outlined in Figure 6.21. Whenever a test failed, the reasons behind the failed test were investigated and the necessary code modifications were made to resolve the issues. The results presented thus far are based on the final revision of the framework implementation which reflects ‘passed’ test runs for all the test cases discussed. However, it is important to discuss various errors that were detected during the development phase that have subsequently been addressed. It is not possible to discuss all the issues here, especially since there have been several minor ‘bug fixes’ ranging from syntax errors to minor logical errors such as attempting to access an out-of-bound element in an array; therefore only the issues that required a considerable debug effort to identify/rectify are discussed here. The important issues, in no particular order are:

i. When identifying medication lapses, a preliminary sequential implementation analysed only the prescriptions that were prescribed after a patient was classified with a required diagnosis. Analysing patient reports indicated that this contained a logical error, since if the patient was classified while on a particular medication, then it would appear as if the patient had no medication at all until the next prescription. Take the example of a patient who was prescribed an antihypertensive for 90 days on 25-Jul-08 (coverage ending on 23-Oct-08). If he/she was then classified with hypertension on 01-Aug-08, and then got prescribed another antihypertensive medication on the 01-Nov-08, the patient has had only a medication lapse of nine days, from 23-Oct-08 to 01-Aug-08. However, if only prescriptions after classification date were included, there would be a 92-day lapse from 01-Aug-08 to 01-Nov-08, which is incorrect. This scenario may not occur in practice as it is unlikely for a patient to be prescribed an antihypertensive and then get classified nine days after, but ChronoMedIt places much emphasises on temporal issues and therefore this was an important case that was resolved.

ii. The SQL Server stored procedure that identifies patients with consecutively high BPs was written so that it used a SQL WHERE clause to identify the high BPs and then checked if the difference between the earliest and latest dates for these BP measurements was greater than the required criterion threshold (120 days for example). However, this was incorrect as the WHERE clause was merely filtering out the controlled BPs that may have been
recorded during the high BPs, so the high BPs that were identified were not necessarily consecutive. The correct way of implementing this was to identify the consecutively high BPs and then filter out the sequences that did not satisfy the minimum number of days between the earliest and the latest BP measurements criterion.

iii. Patients classified before the start of the evaluation period who received their first prescription after the end of the evaluation period should have a lapse duration equivalent to the duration of the evaluation period. These cases were not correctly identified during the initial implementation stages.

iv. During preliminary implementation, ongoing lapses were identified by determining whether the last prescription the patient received was before the end of the evaluation period. Almost always the end of the evaluation period was set to be the data extraction date, however, as evaluation periods prior to the extraction date began to be considered, I realised how it was important to consider the end date of the last prescription that was prescribed before the end of the evaluation period. This process eliminated prescriptions issued after the end of the evaluation period and was used to correctly identify ongoing lapses.

v. Patients were having multiple classifications of the same type on different dates. This issue was raised in Section 3.3.1 and it was decided to use the first classification date when multiple dates were encountered due to the nature of chronic illness.

vi. When evaluation periods prior to the data extraction date were considered, some patients would have prescriptions after the end of the evaluation period. The ongoing lapse durations therefore would often end after the end of the evaluation period. However, this is not the actual lapse duration that needs to be reported by the framework, and therefore the lapse end had to be explicitly set to the end of the evaluation period. The updated lapse duration may or may not satisfy the minimum lapse duration criterion.

vii. It was initially assumed that all prescriptions had valid generic names for drugs. However, during data analysis I found out how some prescriptions contained actual brand names within the generic name field. To account for this issue, commonly used brand names were also included in the drug ontology with OWL:sameAs links used to point to the corresponding generic name. Section 5.5.2 discussed this point in some detail.

6.6 Limitations of Framework Testing

Despite the significant level of testing that has been carried out to ensure reliability of the framework implementation, several limitations need to be addressed:
BVA (and hence robustness testing) has been performed only for single faults. If the system also needs to be tested for multiple faults, then worst-case testing needs to be performed, where the Cartesian product of the five-element set of \( \text{min, min}^+, \text{nom, max}^- \) and \( \text{max} \) (i.e., with reference to the cases in Figure 6.3) forms the set of test cases. Thus, for a function of \( n \) variables, worst-case testing results in \( 5^n \) test cases. Robust worst-case testing is an extension of worst-case testing (see Figure 6.4) where the test cases are identified using the Cartesian product of the seven-element set of \( \text{min}^-, \text{min, min}^+, \text{nom, max}^-, \text{max} \) and \( \text{max}^+ \), resulting in \( 7^n \) test cases [245]. Performing worst-case testing (hence robust worst-case testing) was deemed an effort not worthwhile in the current context since all-pairs testing has been performed on all the equivalence classes, and BVA has been performed on the individual equivalence classes where necessary (for example, BVA has been performed where the lapse period had an overlap with the evaluation period, but not for cases where the lapse ended prior to the evaluation period). However, this means that whether or not a (valid) case with a 31-day lapse at the beginning of evaluation period followed by a second (invalid) lapse of 29 days during the evaluation period is correctly identified as having only one valid lapse occurring at the beginning of the evaluation period has not been tested. The experience with the framework is that these cases are correctly identified as expected, although not explicitly tested using test cases. Further, it is reasonable to assume that such cases would be uncovered by random testing since many such cases would be generated.

Adherence has not been tested systematically although MPR thresholds can be specified in the persistence of medication criterion. Adherence (and therefore MPR) is directly related to persistence and if the lapses can be correctly identified, then it becomes simple math to calculate MPR. However, MPR related criteria with different MPR thresholds have been tested using random testing and therefore it is reasonable to assume that the MPR processing modules have been sufficiently tested.

Random testing has only a certain probability of generating all test cases covering the breadth of the input space and therefore certain values (prescription durations for example) may not have been tested. However, due to the large number of prescriptions generated, this was not considered to be an issue.

There is no guarantee that the sequential implementation (i.e., the golden implementation in Figure 6.20) is 100% defect-free, but this is a central issue surrounding (random) testing. At the initial stages of the sequential-SQL based implementation comparisons, a number of issues were uncovered where the fault was with the sequential based implementation and were subsequently fixed in the C# code. As such, the sequential code base has gradually evolved into a robust framework, and with the random testing that has been performed...
with the large number of instances, it is reasonable to assume that the sequential
implementation is a defect-free implementation although this is not guaranteed to be so.
• Audit criteria testing has been carried out only for the cases shown in Table 6.17. It is
possible to test medication lapses and/or MPR in different types of medications other than
the ones specified in Table 6.17, for example, lapses in diuretics. Testing for all possible
drugs or drug classes is not feasible due to the effort required, and it is believed that the
cases discussed here are sufficiently representative of all drugs and drug classes.
Chapter 7

Applications using ChronoMedIt

The previous chapters discussed the various features of ChronoMedIt and its implementation details. The purpose of this chapter is to demonstrate several applications of ChronoMedIt and discuss some studies/data analyses that have been carried out using the framework. Each study is associated with a different aspect of quality audit reporting for quality improvement purposes. The studies this chapter covers are related to (1) determining patients satisfying the eight audit criteria; (2) investigating the relationship between good ACEi/ARB adherence and BP control in patients with diabetes; (3) investigating the relationship between good antihypertensive adherence and BP control in patients with hypertension; (4) investigating the use of interval based quality measures opposed to point-in-time measures; (5) discussing how two of the selected indicators from the QOF can be improved; (6) comparing dispensing based adherence to prescribing based adherence; and (7) demonstrating the use of the framework to determine medication adherence in a different domain. These studies are followed by a brief discussion on an ongoing nurse-led feasibility trial for antihypertensive medication adherence promotion.

7.1 Data Extract

My supervisor and I were collaborating with two Auckland based general practices, and we extracted patient EMR data from the commercial PMS of these two practices for the 24-month period from 01-Apr-07 to 31-Mar-09 (with the exception of classifications which were extracted for as far back as possible). The data extraction protocol was similar to that mentioned in
Section 3.2 and only funded patients enrolled at the practices were included. Details of the extracted dataset are shown in Table 7.1, and most of the analyses discussed in this chapter are based on this dataset unless specified otherwise. The study protocol was approved under the NZ Multiregional Ethics Committee protocol number MEC/09/32/EXP.

Table 7.1: Summary of the two practice datasets for funded and enrolled patients

<table>
<thead>
<tr>
<th></th>
<th>Practice-1</th>
<th>Practice-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Funded patients</td>
<td>5454</td>
<td>4424</td>
</tr>
<tr>
<td>Patients with classifications</td>
<td>4576</td>
<td>3855</td>
</tr>
<tr>
<td>Hypertension</td>
<td>602</td>
<td>511</td>
</tr>
<tr>
<td>Diabetes</td>
<td>514</td>
<td>419</td>
</tr>
<tr>
<td>Renal impairment</td>
<td>146</td>
<td>123</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>27</td>
<td>31</td>
</tr>
<tr>
<td>BP measurements</td>
<td>11977</td>
<td>11637</td>
</tr>
<tr>
<td>Prescriptions</td>
<td>61840</td>
<td>54353</td>
</tr>
<tr>
<td>Antihypertensives</td>
<td>7663</td>
<td>6421</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 30</td>
<td>2949</td>
<td>2383</td>
</tr>
<tr>
<td>30-44</td>
<td>1149</td>
<td>826</td>
</tr>
<tr>
<td>45-59</td>
<td>838</td>
<td>706</td>
</tr>
<tr>
<td>60-74</td>
<td>399</td>
<td>396</td>
</tr>
<tr>
<td>75+</td>
<td>119</td>
<td>113</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>2748</td>
<td>2216</td>
</tr>
<tr>
<td>Male</td>
<td>2706</td>
<td>2208</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maori</td>
<td>459</td>
<td>301</td>
</tr>
<tr>
<td>Pacific Island</td>
<td>4140</td>
<td>3970</td>
</tr>
<tr>
<td>European</td>
<td>154</td>
<td>41</td>
</tr>
<tr>
<td>Asian</td>
<td>161</td>
<td>32</td>
</tr>
<tr>
<td>Other</td>
<td>540</td>
<td>80</td>
</tr>
</tbody>
</table>

7.2 Identifying patients satisfying the eight audit criteria

For this study, the evaluation period was defined as the one year period from 01-Apr-08 to 31-Mar-09 with a six month run-in period. Using the various attributes from the Criteria Model (as demonstrated in Table 5.3), the eight criteria were formulated, and the resulting numbers of patients are shown in Table 7.2.
Table 7.2: Number of patients classified with hypertension satisfying the eight audit criteria. EP refers to the evaluation period.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Patients Satisfying Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practice-1 (N = 602)</td>
</tr>
<tr>
<td>C1  A lapse in antihypertensive therapy &gt;30 days and the lapse extends into the EP</td>
<td>340 (56%)</td>
</tr>
<tr>
<td>C2  A period of &gt;180 days with no BP measurements extending into the EP</td>
<td>248 (41%)</td>
</tr>
<tr>
<td>C3  A BP measurement of $\geq 160/100$ mmHg followed by a gap of &gt;120 days in BP measurements extending into the EP</td>
<td>34 (6%)</td>
</tr>
<tr>
<td>C4  Three or more consistently high BP measurements ($\geq 160/100$ mmHg) over 120 days or more where either i) the last of these high BPs was within the EP or ii) with no subsequent “controlled” BP ($&lt; 160/100$ mmHg) measurements after the consistently high BPs</td>
<td>6 (1%)</td>
</tr>
<tr>
<td>C5  Classified with diabetes mellitus and not on ACEi/ARB at any time during EP</td>
<td>238 (40%)</td>
</tr>
<tr>
<td>C6  Classified with myocardial infarction and not on beta-blocker at any time during EP</td>
<td>14 (2%)</td>
</tr>
<tr>
<td>C7  Classified with renal impairment and on ACEi/ARB and with eGFR &lt; 60mL/min at any time during EP</td>
<td>40 (7%)</td>
</tr>
<tr>
<td>C8  On thiazide(s) and with serum uric acid $&gt; 0.42$mmo/l at any time during EP</td>
<td>68 (11%)</td>
</tr>
</tbody>
</table>

These results based on production EMR data indicate that a significant portion of patients with hypertension, ranging from 56%-63%, have >30 day lapses in their antihypertensive medication. Over a third of people with hypertension have not had a BP measurement for >180 days while a considerable proportion of patients (40% for both practices) with hypertension and diabetes had lapses in their ACEi/ARB medication. The patient numbers in Table 7.2, especially with respect to C1, C2 and C5 illustrate the significant opportunity for intervention to improve the chronic disease management process of these patients and thereby improve patient outcomes.

7.3 Investigating ACEi/ARB adherence issues and the relationship to BP control in patients with hypertension and diabetes

Treating with an ACEi/ARB is an important and indicated component of most regimens to control BP in hypertensive patients with diabetes as investigated by other researchers [6, 269].
The American Diabetes Association recommends ACEi/ARB as a first-line therapy for the treatment of hypertension in persons with diabetes due to the ability of these drugs to slow the development and/or progression of diabetic nephropathy (i.e., kidney disease due to diabetes) [270]. To identify hypertensive patients with diabetes who have poor adherence to ACEi/ARB, the ChronoMedIt framework was used under three specific scenarios: i) ACEi/ARB MPR less than 80%, ii) a lapse in ACEi/ARB for greater than 30 days and iii) ACEi/ARB MPR less than 80% with a lapse greater than 30 days. A one year evaluation period from 01-Oct-07 to 30-Sep-08 with a six month run-in period was used for analysis.

Only patients under active ACEi and/or ARB management were included in this study - i.e., patients with at least one ACEi/ARB prescription during the evaluation period or during the run-in period and extending coverage into the evaluation period (hence MPR >0). The resulting patient numbers are shown in Table 7.3.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Patients Satisfying Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practice-1 (N = 244)</td>
</tr>
<tr>
<td>1 Non-zero ACEi/ARB MPR &lt;80%</td>
<td>85 (56%)</td>
</tr>
<tr>
<td>2 Lapse in ACEi/ARB for greater than 30 days</td>
<td>117 (7%)</td>
</tr>
<tr>
<td>3 Non-zero ACEi/ARB MPR less than 80% with a lapse in ACEi/ARB for greater than 30 days</td>
<td>85 (11%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Patients Satisfying Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practice-2 (N = 163)</td>
</tr>
<tr>
<td>1 Non-zero ACEi/ARB MPR &lt;80%</td>
<td>71 (63%)</td>
</tr>
<tr>
<td>2 Lapse in ACEi/ARB for greater than 30 days</td>
<td>96 (8%)</td>
</tr>
<tr>
<td>3 Non-zero ACEi/ARB MPR less than 80% with a lapse in ACEi/ARB for greater than 30 days</td>
<td>70 (18%)</td>
</tr>
</tbody>
</table>

To determine the relevance of the reporting to patient outcomes, the relationship of MPR was analysed, as yielded by the computational framework, to BP control. JNC7 [6] recommends a goal BP of <130/80 mmHg for patients with hypertension and diabetes, and as such, ‘high BP’ for the purposes of this study was defined as having systolic BP ≥130 mmHg or diastolic BP ≥80 mmHg (the NZ Heart Foundation guidelines [271] use a more sophisticated BP target that slides according to overall cardiovascular risk). Using the earliest BP measurement during the six-month period after the end of evaluation period, that is between 01-Oct-08 – 31-March-09 as the outcome measure, a 2x2 table was constructed as per Table 7.4 for patients with hypertension and diabetes under active ACEi/ARB management.
Table 7.4: ACEi/ARB adherence vs. BP in patients with hypertension and diabetes. N represents the number of active patients with a BP measurement after the end of the evaluation period.

<table>
<thead>
<tr>
<th></th>
<th>Practice-1 (N = 195)</th>
<th>Practice-2 (N = 146)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High BP</strong></td>
<td><strong>BP Controlled</strong></td>
<td><strong>High BP</strong></td>
</tr>
<tr>
<td>MPR &lt;80%</td>
<td>66</td>
<td>6</td>
</tr>
<tr>
<td>MPR ≥80%</td>
<td>90</td>
<td>33</td>
</tr>
<tr>
<td><strong>Odds Ratio (OR)</strong></td>
<td>4.0</td>
<td>(p=0.002, 95% CI 1.6-9.9)</td>
</tr>
</tbody>
</table>

The results in Table 7.4 indicate that non-adherent patients with hypertension and diabetes are at least two times more likely to have uncontrolled BP than adherent patients (OR = 4.0, p = 0.002 and OR = 2.5, p = 0.034 for the two practices). The result is unsurprising, per se, and agrees with past findings about the impact of non-adherence on BP control [272]. However, the result validates that analysing electronic prescribing (as compared to dispensing, or some other more proximate source of data) is important to provide a significant indication of poor BP control. Further, the high odds ratios observed indicate the importance of medication adherence (for having controlled BP) and warrants active adherence promotion.

7.4 Antihypertensive adherence and impact on BP control

The main purpose of this study was to determine how good a predictor antihypertensive MPR is for having controlled BP. The same evaluation period and run-in period used in Section 7.3 were used for this study as well. The main inclusion criteria for this analysis were for the patients to have a hypertension diagnosis, be on active antihypertensive therapy (i.e., MPR >0) and have a BP measurement during the six-months after the end of the evaluation period. ‘High BP’ was defined as having a systolic BP ≥140 mmHg or diastolic BP ≥90 mmHg, and systolic BP ≥130 mmHg or diastolic BP ≥80 mmHg if the patient also had a diabetes classification.

Table 7.5: Antihypertensive adherence vs. BP in patients with hypertension. N represents the number of active patients with a BP measurement after the end of the EP

<table>
<thead>
<tr>
<th></th>
<th>Practice-1 (N = 355)</th>
<th>Practice-2 (N = 334)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High BP</strong></td>
<td><strong>BP Controlled</strong></td>
<td><strong>High BP</strong></td>
</tr>
<tr>
<td>MPR &lt;80%</td>
<td>98</td>
<td>24</td>
</tr>
<tr>
<td>MPR ≥80%</td>
<td>147</td>
<td>86</td>
</tr>
<tr>
<td><strong>Odds Ratio (OR)</strong></td>
<td>2.4</td>
<td>(p=0.001, 95% CI 1.4-4.0)</td>
</tr>
</tbody>
</table>
The odds ratios in Table 7.5 indicate a reasonably good association between MPR for antihypertensives and having controlled BP. The charts in Figure 7.1 and Figure 7.2 demonstrate how average systolic BP is correlated to antihypertensive MPR for the two practices. Using a linear regression model, for Practice-1 the regression coefficient for MPR is 0.1931 while the intercept is 154.63 (model $R^2=0.0983$); for Practice-2, the regression coefficient for MPR is 0.1639 and the intercept is 148.8 (model $R^2=0.0616$). The observed mean reductions in systolic BP for being adherent from 0% to 100% are 19.31 mmHg and 16.39 mmHg respectively for the two practices.

Figure 7.1: Plot of systolic BP against antihypertensive MPR for patients with hypertension for Practice-1
Similar to the previous study that showed that good ACEi/ARB adherence is important for patients with hypertension and diabetes to have controlled BP, this study indicates that good antihypertensive adherence is important for patients with hypertension to have controlled BP.

7.5 Interval Based Measures as Quality Indicators in Blood Pressure Management

As discussed in Chapter 2 (primarily Section 2.4), the medical community in general accepts that following well developed, evidence-based clinical practice guidelines improves the quality of care received by patients along with the many other benefits guidelines provide [40]. Together with guidelines, quality indicators are often used in order to provide feedback to clinicians and give an indication of the quality of patient care delivered. The QOF in the UK [196] provides a set of clinical indicators across four domains (clinical, organisational, additional services and patient experience) designed around best practice to improve the quality of service provided to patients. Each indicator is allocated a number of points and GP practices are awarded points according to how well they have performed, with associated monetary compensation.

It is interesting to note that most of the widely used quality indicators use the presence of a single point-in-time measurement to determine whether a given indicator is satisfied. For example, an important QOF indicator with the highest point allocation (57 QOF points) in the
‘clinical domain’ is BP5 – “The percentage of patients with hypertension in whom the last blood pressure (measured in the previous nine months) is 150/90 or less” [196]. This uses the notion of the ‘last BP’ being controlled. The eight quality indicators that have been discussed throughout this thesis have a strong association to temporal intervals (i.e., to the evaluation period), rather than focusing on a single measurement at one point in time. The QOF’s BP5 suggests that its framers saw relevance to nine months as an appropriate evaluation period (and, in this sense BP5 is at least partially interval based). For this particular study, five BP quality indicators are examined ranging from a completely point-in-time ‘last BP high’ measure to a range of more interval-oriented indicators. Their stability (ability to predict the same indicator) over time and their association with each other are also investigated.

The two indicators (with modified thresholds for comparability) out of the eight indicators that are relevant to QOF’s BP5 are:

1. “patients classified with hypertension with two or more consistently high BP measurements (>150/90 mmHg) over 90 days or more where either (i) the last of these high BPs was within the evaluation period or (ii) with no subsequently ‘controlled’ BP (≤150/90 mmHg) measurement after the consistently high BPs” and;
2. “patients classified with hypertension with a BP measurement >150/90 mmHg followed by a gap of >120 days in BP measurements extending into the evaluation period”

These indicators are referred to as ‘Consistent high BP’ and ‘High BP then lapse’ indicator respectively. The ‘Evaluation period’ refers to the period of interest; in the context of BP5, it will be nine months (since the QOF specifies that the BPs relevant to the BP5 indicator should have been measured during the previous nine months). A six-month ‘run-in’ period was also used to account for intervals of suboptimal management that begin before the evaluation period but extend into it.

7.5.1 Analysis Protocol

Failure rates were analysed as of the end of the first quarter of 2009 using five different quality indicators – (1) last BP high (>150/90 mmHg), (2) failing BP5, (3) consistently high BP, (4) high BP then lapse and (5) MPR <80%. Failing BP5 in fact consists of two patient cohorts – patients with BPs where the last BP was not controlled (i.e., explicitly high, equivalent to ‘last BP high’), and patients who do not have a BP measurement during the nine-month period. Using the five quality indicators, one can determine prediction rates at one quarter, two quarters and three quarters of a year prior to the evaluation date to determine the suitability of each indicator to
accurately predict patients who will continue to fail each aspect of BP management. The dates of interest are illustrated in Figure 7.3.

![Figure 7.3: Timelines related to the quality indicator analysis](image)

Using the timelines in Figure 7.3, the various rates of failure (i.e., the patient having the specific deficiency in GP management) were computed as predicted by the same indicator three-months, six-months and nine-months prior to 31-Mar-09. In terms of conditional probability, the probability that a patient will fail an indicator given that the patient failed at a prior date is given by:

\[
p(t \mid (t - k)) = \frac{p(t) \cap p(t - k)}{p(t - k)}
\]

where \(p(t)\) indicates the probability that a patient will fail the indicator during the current evaluation period while \(p(t - k)\) indicates the probability that a patient will fail the indicator during a time period prior to the evaluation period (with \(k\) of three-months, six-months and nine-months). This conditional probability is equivalent to the PPV of the indicator at time \(t - k\) for the indicator at time \(t\).

To assess the degree to which the indicators agree with one another (and hence are potentially ‘redundant’) the conditional probability was looked at of a patient failing one indicator given that they have failed another with respect to the most current evaluation period. Again, this conditional probability is equivalent to PPV.
Note that for the MPR criterion there is an additional inclusion constraint that a patient is required to have at least one antihypertensive prescription during the 15-month period (i.e., during the run-in period or the evaluation period).

### 7.5.2 Data analysis

535 and 462 patients from Practice-1 and Practice-2, respectively, were funded and enrolled at 1-Jul-08 and had a hypertension diagnosis; of these, 459 (86%) and 397 (86%) patients, from Practice-1 and Practice-2 respectively, satisfied the additional inclusion criterion for computation of MPR. The predictive abilities of the five indicators are shown in Table 7.6.

| Table 7.6: PPVs of quality indicators based on past performance on same indicator. Denominator is patients failing in evaluation period ending at time \( t - k \); numerator is patients failing on ‘current’ evaluation period (ending at time \( t \)) and also failing in evaluation period ending at time \( t - k \). |
|---|---|---|---|---|---|
| \( k \) | Practice | Last BP high | Failed BP5 | Consistent high BP | High BP then lapse | MPR <80% |
| 3-months | Practice-1 | 65/118 = 55% | 133/190 = 70% | 29/39 = 74% | 94/118 = 80% | 143/166 = 86% |
| Practice-2 | 62/107 = 58% | 104/152 = 68% | 30/35 = 86% | 48/67 = 72% | 141/176 = 80% |
| 6-months | Practice-1 | 56/142 = 39% | 116/198 = 59% | 27/51 = 53% | 63/101 = 62% | 133/184 = 72% |
| Practice-2 | 42/95 = 44% | 92/145 = 64% | 23/34 = 68% | 38/69 = 55% | 120/173 = 69% |
| 9-months | Practice-1 | 49/156 = 31% | 108/210 = 51% | 22/45 = 49% | 38/82 = 46% | 121/188 = 64% |
| Practice-2 | 34/96 = 35% | 81/149 = 54% | 14/27 = 52% | 29/67 = 43% | 114/182 = 63% |

These results show that ‘last BP high’, a point-in-time measure, is less stable than the other, interval-oriented, indicators. Failing BP5 ranks reasonably well; its PPV is generally under that of consistent high BP, high BP then lapse and MPR<80% for the three-month and six-month prior intervals, but this owes in part to the interval measures being correlated by definition when the time periods overlap. MPR<80% is the most stable criterion among the five indicators considered.

Table 7.7 and Table 7.8 show the PPVs between quality indicators for the two practices at time \( t \); for those PPVs not definitionally 100% the median is 43% (min = 10%, max = 79%).
Table 7.7: PPVs between different indicators for Practice-1. The values indicate the number of patients failing an indicator given another indicator failed by \( n \) patients, followed by the corresponding probability (as a percentage).

<table>
<thead>
<tr>
<th>Given</th>
<th>Last BP high</th>
<th>Failed BP5</th>
<th>Consistent high BP</th>
<th>High BP then lapse</th>
<th>MPR &lt;80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last BP high (( n=99 ))</td>
<td>99, 100%</td>
<td>99, 100%</td>
<td>21, 21%</td>
<td>47, 47%</td>
<td>50, 51%</td>
</tr>
<tr>
<td>Failed BP5 (( n=179 ))</td>
<td>99, 55%</td>
<td>179, 100%</td>
<td>22, 12%</td>
<td>60, 34%</td>
<td>77, 43%</td>
</tr>
<tr>
<td>Consistent high BP (( n=38 ))</td>
<td>21, 55%</td>
<td>22, 58%</td>
<td>38, 100%</td>
<td>22, 58%</td>
<td>17, 45%</td>
</tr>
<tr>
<td>High BP then lapse (( n=109 ))</td>
<td>47, 43%</td>
<td>60, 55%</td>
<td>22, 20%</td>
<td>109, 100%</td>
<td>64, 59%</td>
</tr>
<tr>
<td>MPR &lt;80% (( n=163 ))</td>
<td>50, 31%</td>
<td>77, 47%</td>
<td>17, 10%</td>
<td>64, 39%</td>
<td>163, 100%</td>
</tr>
</tbody>
</table>

Table 7.8: PPVs between different indicators for Practice-2. The values indicate the number of patients failing an indicator given another indicator failed by \( n \) patients, followed by the corresponding probability (as a percentage).

<table>
<thead>
<tr>
<th>Given</th>
<th>Last BP high</th>
<th>Failed BP5</th>
<th>Consistent high BP</th>
<th>High BP then lapse</th>
<th>MPR &lt;80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last BP high (( n=102 ))</td>
<td>102, 100%</td>
<td>102, 100%</td>
<td>31, 30%</td>
<td>33, 32%</td>
<td>51, 50%</td>
</tr>
<tr>
<td>Failed BP5 (( n=158 ))</td>
<td>102, 65%</td>
<td>158, 100%</td>
<td>31, 20%</td>
<td>44, 28%</td>
<td>67, 42%</td>
</tr>
<tr>
<td>Consistent high BP (( n=39 ))</td>
<td>31, 79%</td>
<td>31, 79%</td>
<td>39, 100%</td>
<td>16, 41%</td>
<td>20, 51%</td>
</tr>
<tr>
<td>High BP then lapse (( n=62 ))</td>
<td>33, 53%</td>
<td>44, 71%</td>
<td>16, 26%</td>
<td>62, 100%</td>
<td>40, 65%</td>
</tr>
<tr>
<td>MPR &lt;80% (( n=168 ))</td>
<td>51, 30%</td>
<td>67, 40%</td>
<td>20, 12%</td>
<td>40, 24%</td>
<td>168, 100%</td>
</tr>
</tbody>
</table>

7.5.3 Visualisation

Using the graphical tool developed as part of the ChronoMedIt framework, Figure 7.4 illustrates a selected patient’s prescribing patterns together with physiological outcomes.
The patient shown in Figure 7.4 has been classified with hypertension and has failed all five indicators. The last BP is high and therefore satisfies the ‘last BP high’ and ‘failed BP5’ indicators. There are two or more consistently high BPs as all three measurements are high with the difference between first and last high BPs being 386 days. Two instances of high BP then lapse can be observed in Figure 7.4 - (1) 218 days between first and second BP measurements; and (2) 168 days between second and third BP measurements. Also, there are two lapses in antihypertensive therapy (a lapse of 42 days from 26-Aug-08 to 7-Oct-08 and of 78 days from 5-Jan-09 to 24-Mar-09); thus there is a total of 120 days without medication for a 273 day evaluation period, and MPR = (273-120) / 273 = 56.04%, therefore the patient fails the MPR<80% indicator.

### 7.5.4 Interpretation

Results of this study indicate that MPR <80% is the most stable measure, and that all interval based measures are more stable than the point-in-time measure of ‘last BP high’. All the indicators demonstrated herein are important for actively identifying and managing patients on suboptimal BP control and the low PPVs shown in Table 7.7 and Table 7.8 indicate that relying on a single measure is not sufficient as the cohorts identified by each indicator are considerably different.
Table 7.6 shows at least 74% of the patients from both practices who failed consistent high BP, high BP then lapse and MPR at the end of 2008 continued to fail these criteria at the end of the first quarter in 2009, demonstrating detectable opportunities for improved case management.

7.6 Comparison to two indicators from the Quality and Outcomes Framework to Enhance Quality of Pay-for-Performance Incentives

Two important QOF indicators with the highest point allocations under ‘Hypertension’ and ‘Diabetes mellitus’ are BP5 – “The percentage of patients with hypertension in whom the last blood pressure (measured in the previous 9 months) is 150/90 or less” (i.e., the same indicator used for the study in Section 7.5) and DM12 – “The percentage of patients with diabetes in whom the last blood pressure (measured in the previous 15 months) is 145/85 or less” with 57 and 18 points respectively [196]. The previous study investigated the relevance of different interval-based indicators as relevant to BP5, while this study investigates how satisfying BP5 or DM12 is not necessarily an indication that a patient has been on optimal therapy. Three time interval based indicators are considered for demonstration purposes and as relevant in the present context (with modified thresholds to match BP5) these indicators are:

1. “patients classified with hypertension with two or more consistently high BP measurements (>150/90 mmHg) over 90 days or more where either (i) the last of these high BPs was within the evaluation period or (ii) with no subsequently ‘controlled’ BP (≤150/90 mmHg) measurement after the consistently high BPs” and;
2. “patients classified with hypertension with a BP measurement of >150/90 mmHg followed by a gap of >120 days in BP measurements extending into the evaluation period”
3. “patients classified with hypertension with a lapse in BP measurement >180 days extending into the evaluation period”

These indicators are referred to as the ‘consistently high BP’ indicator, the ‘high BP-then-lapse’ indicator and the ‘no BP measurement’ indicator, respectively. Similar to BP5, within the context of DM12, the above indicators refer to patients classified with diabetes (instead of hypertension) with a BP threshold of 145/85 mmHg (instead of 150/90 mmHg), and the evaluation period is 15 months (to mirror the BP timeframe of DM12).

Table 7.9 and Table 7.10 show the numbers of patients with hypertension and diabetes, respectively, satisfying each of their relevant quality indicators.
Table 7.9: Patients with hypertension satisfying different quality indicators

<table>
<thead>
<tr>
<th></th>
<th>Practice-1 (N = 602)</th>
<th>Practice-2 (N = 511)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failing BP5</td>
<td>206 (34%)</td>
<td>181 (35%)</td>
</tr>
<tr>
<td>‘Consistently high BP’</td>
<td>38 (6%)</td>
<td>43 (8%)</td>
</tr>
<tr>
<td>‘High BP-then-lapse’</td>
<td>111 (18%)</td>
<td>66 (13%)</td>
</tr>
<tr>
<td>‘No BP measurement’</td>
<td>236 (39%)</td>
<td>178 (35%)</td>
</tr>
</tbody>
</table>

Table 7.10: Patients with diabetes satisfying different quality indicators

<table>
<thead>
<tr>
<th></th>
<th>Practice-1 (N = 514)</th>
<th>Practice-2 (N = 419)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failing DM12</td>
<td>196 (38%)</td>
<td>173 (41%)</td>
</tr>
<tr>
<td>‘Consistently high BP’</td>
<td>57 (11%)</td>
<td>30 (7%)</td>
</tr>
<tr>
<td>‘High BP-then-lapse’</td>
<td>147 (29%)</td>
<td>70 (17%)</td>
</tr>
<tr>
<td>‘No BP measurement’</td>
<td>260 (51%)</td>
<td>144 (34%)</td>
</tr>
</tbody>
</table>

Table 7.11 shows the number of patients who satisfied their relevant QOF indicator, but also failed any of the three interval based indicators. Totalling across both the practices, 34% (95% CI: 30% - 37%) of hypertension patients satisfying BP5 fail at least one of the interval based indicators and 44% (95% CI: 40% - 48%) of the diabetes patients satisfying DM12 fail at least one of the interval based indicators.

Table 7.11: Patients satisfying a QOF indicator (BP5/DM12), but failing any one of the three indicators of interest

<table>
<thead>
<tr>
<th></th>
<th>Practice-1</th>
<th>Practice-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Consistently high BP’ or ‘High BP-then-lapse’ or ‘No BP measurement’, given BP5</td>
<td>149/396 = 38%</td>
<td>96/330 = 29%</td>
</tr>
<tr>
<td>‘Consistently high BP’ or ‘High BP-then-lapse’ or ‘No BP measurement’, given DM12</td>
<td>172/318 = 54%</td>
<td>76/246 = 31%</td>
</tr>
</tbody>
</table>

7.6.1 Comparing consistently high BP with BP5 and DM12

A patient having consistently high BPs is an indication of ‘taking too long’ to reach target BP levels and therefore these patients need to be actively managed. To fail this indicator, it is required to have at least two uncontrolled BP measurements occurring over 90 days with either (i) the last of these high BPs measured during the evaluation period or (ii) with no subsequently ‘controlled’ BP measurement after the consistently high BPs (the last high BP would occur during the run-in period in this case). If a particular patient satisfied this indicator due to scenario (ii) it means that the patient has not had BP measured during the entire evaluation
period. These cases are important as these patients have had consistently high BPs, but their BP has not been measured since. With respect to diabetes patients, the NICE guideline (recommendation R64 [273]) suggests monitoring BP 1–2 monthly, and intensifying therapy if on medications, until BP is consistently below 140/80 mmHg; therefore failing this indicator (despite the slight variation in BP thresholds) does not meet guideline recommendations as well. Considering data from Table 7.9 and Table 7.10 indicate that prevalence of consistently high BP (in the sense of the interval based indicator, i.e., repeatedly measured and recorded) is not very high and ≤8% for the hypertension cohort and ≤11% for the diabetes cohort for a given practice.

7.6.2 Comparing high BP then lapse with BP5 and DM12

A patient having a high BP and then a long lapse in measurement is an indication of ‘taking too long to check whether the patient has reached target BP’. This has been particularly documented in the QOF justification for DM12 which suggests that high BP in people with diabetes should be treated aggressively with lifestyle modification and drug therapy [196] and also in the NICE guidelines (recommendation R61 [273]) where repeated BP measurements are suggested every two months if BP is higher than 140/80 mmHg. Therefore, diabetic patients with an uncontrolled BP who have not had their BP measured for a significant period of time (at least 120 days in our case) should be actively recalled and managed by the practice. Actively managing patients can be viewed as a more ‘aggressive’ approach as opposed to evaluating a diabetic patient’s last BP. With respect to this indicator, a patient with hypertension may have had an uncontrolled BP measure just under nine months prior to end of the evaluation period and just under 15 months in the case of a diabetic patient, and a second ‘controlled’ BP towards the end of the evaluation period, thereby also satisfying BP5 or DM12; however, this type of management may not necessarily demonstrate optimal BP control worthy of an incentive for a practice. Table 7.9 and Table 7.10 indicate that prevalence of high BP then lapse is higher than having consistently high BPs (≤18% for the hypertension cohort and ≤29% for the diabetes cohort for a given practice).

7.6.3 Comparing no BP measurement with BP5 and DM12

The NICE guidelines recommend (recommendation R71 [273]) monitoring BP of a person who has attained and consistently remained at their BP target every 4–6 months, and checking for possible adverse effects of antihypertensive therapy – including the risks from unnecessarily low BP. If ‘no BP measurement’ is taken, this recommendation is not achieved. Identifying these
patients would allow proactive recall and management. Compared to the other two indicators, this has the highest prevalence in the two practices, ranging from 34% to 51%.

7.6.4 Interpretation

Having a controlled last BP measurement is naturally an important indicator; however, this study indicates that there is also room for practices to actively identify patients in whom BP is not measured or controlled in a timely fashion, and that this is not entirely captured by final BP alone. In particular, gaps were identified on the interval based measures on the order of 34% for hypertension patients satisfying BP5, and on the order of 44% for diabetes patients satisfying DM12.

The results in Table 7.9 and Table 7.10 indicate that the two practices considered in this study would perform well on BP5 and very well on DM12 which need a 40-70% range and a 40-60% range respectively to receive the maximum QOF points, and thereby incentives for managing the patients well. The other indicators considered herein (other than BP5 and DM12) are important, and patients satisfying these indicators suggests that they may not necessarily be on optimal therapy, despite the practices receiving BP5 and/or DM12 incentives for managing these patients well.

7.7 Comparing dispensing based adherence to prescribing based adherence

In NZ, the government pays about 67% of overall pharmaceutical expenses [274] as determined by a Pharmaceutical Schedule set by the Pharmaceutical Management Agency (PHARMAC). Community pharmacies submit electronic claims for reimbursement of the government funded components of dispensed drugs to a centralised reimbursement system. NZ has a very computer-literate pharmacy sector, with more than 98% of pharmacies using computers as part of their business [275]. In addition, general medical practice in NZ has a near 100% computerisation rate [7] with practices using one of several proprietary PMSs. The PMS maintains electronic records of prescribing, although the prescription is then printed to paper for the patient to take to the pharmacy. NZ has a well established NHI number; and general practices are financially rewarded through their PHOs for maintaining accurate registers of their patients’ NHIs [193].

Within this healthcare setting, there is the opportunity to examine the information value of general practice electronic prescribing records with respect to individual adherence to common
long-term medications that feature heavily in chronic condition management. These features of the NZ healthcare system allow one to compare general practice prescribing to community based dispensing with respect to what each says about adherence of individuals to long-term medication, and, in particular, to assess the extent to which measures of adherence based on prescribing predict adherence as measured by dispensing.

In this study, the prescribing and dispensing of six generic drugs are examined, each of which is listed on the Pharmaceutical Schedule and is therefore fully or partly funded by government, is (based on analysis of general practice prescribing records) the most commonly prescribed of its class, and which serves a key role in management of common chronic conditions: (1) simvastatin to represent statins, (2) metoprolol succinate to represent beta blockers, (3) bendrofluazide for thiazide diuretics, (4) cilazapril for ACEi, (5) felodipine for dihydropyridine calcium channel blockers and (6) metformin for oral hypoglycaemic agents.

### 7.7.1 Data Extraction and Matching

The data extraction and matching process was approved under the Northern Regional Ethics Committee protocol number NTX/07/55/EXP. Note that this dataset was different from the ones used in the previous studies, although the data extraction process from the PMS was similar.

For the purposes of this study, my supervisor and I collaborated with a multi-physician NZ general medical practice in the south Auckland suburbs. Prescribing data (along with patient gender, age, ethnicity and socio-economic coding) was extracted from this practice’s proprietary PMS (MedTech32) into a password-protected research database, retaining NHI, but removing other identifying information including name and address. The pharmacy-dispensing records for government subsidised medicines for these patients were extracted from the national claims database, as matched on NHI, by the NZ Health Information Service (now part of the Information Directorate) of the Ministry of Health, and then merged into the research database. Data were analysed by non-proprietary, generic, active ingredient in order to account for brand and generic substitution.

The same persistence based adherence measures of proportion-of-days-covered was employed for this analysis based on the prescribing records in the PMS of the patient’s general practice, which is denoted by MPR_Pr. For the purpose of this study MPR_Dp is used to refer to the proportion-of-days-covered by medication dispensed from community pharmacies, as
calculated from the dispensing records in the national claims database. Similar to previous MPR analyses in this thesis, an MPR less than 80% was used as a threshold upon which clinical effectiveness of a therapy is significantly compromised.

### 7.7.2 Analysis Protocol

The prescribing of the six selected drugs was examined to funded patients enrolled at the practice and the associated community pharmacy dispensing with respect to the evaluation period 01-Jan-06 to 30-Mar-07. Rates of high adherence were then examined, for each drug and overall, as indicated on MPR_Pr and MPR_Dp; and the predictive value of MPR_Pr for MPR_Dp, in terms of the probability of non-adherence in MPR_Pr indicating non-adherence in MPR_Dp. Correspondence of a prescribing record to a timely (within seven days) dispensing record for the same drug, and linear regression for MPR_Dp incorporating demographic and socioeconomic factors along with MPR_Pr were also calculated.

### 7.7.3 Prescribing-Dispensing dataset

The data extracted from the PMS contained prescribing records for the 18-month period from 01-Oct-05 to 30-Mar-07. The data extract included 2713 patient records and 29772 prescribing records for 636 different generic drugs as recorded by the PMS. The pharmacy dataset included dispensing data for the 21-month period from 01-Oct-05 to 30-Jun-07 for these 2713 patients, linked by their NHI. In NZ, a prescription (with refills) is valid for a period of three months before it expires (medication may be dispensed as three months supply at once, or may be collected one month at a time from a single three-month prescription); hence the pharmacy dispensing data was extracted for an additional three months beyond the date of the last prescription. The dataset included data for 2713 patients who had 63833 dispensing records for 410 different generic drugs from 170 different pharmacies in NZ. The pharmacy dataset includes fewer generic drugs, because not all drugs are subsidised.

### 7.7.4 Data analysis

A 15-month evaluation period from 01-Jan-06 to 30-Mar-07 was considered with a three-month run-in period. A total of 646 patients were prescribed at least one of the six selected long-term medications either during the evaluation period or with a supply extending into the evaluation period. These cases form the cohort for subsequent analysis. In addition, 117 of the funded patients have dispensing records for one or more of the six drugs but no prescribing records for those drugs in the practice PMS – these cases (presumably managed by other physicians) do not form part of the further analysis.
7.7.5 Prescribed and Dispensed Drugs

The prescribing frequencies for the six generic drugs of interest for the cohort of 646 patients are shown in Table 7.12.

<table>
<thead>
<tr>
<th>Generic Drug Name</th>
<th>Number of Prescriptions</th>
<th>Number of Dispensings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Simvastatin</td>
<td>1159</td>
<td>2928</td>
</tr>
<tr>
<td>2. Metoprolol succinate</td>
<td>949</td>
<td>2569</td>
</tr>
<tr>
<td>3. Bendrofluazide</td>
<td>624</td>
<td>1367</td>
</tr>
<tr>
<td>4. Cilazapril</td>
<td>566</td>
<td>1750</td>
</tr>
<tr>
<td>5. Felodipine</td>
<td>426</td>
<td>973</td>
</tr>
<tr>
<td>6. Metformin</td>
<td>319</td>
<td>899</td>
</tr>
</tbody>
</table>

Overall, the frequencies of prescribing and dispensing follow a similar pattern. The number of dispensing records approaches three times the number of prescriptions, which is consistent with most long-term prescriptions being issued with two repeats/refills.

7.7.6 MPR_Pr vs. MPR_Dp

For each of the six drugs, the MPR_Pr and MPR_Dp were for the evaluation period 01-Jan-06 to 30-Mar-07. Table 7.13 shows rates of high adherence based on MPR_Pr and MPR_Dp by drug.

<table>
<thead>
<tr>
<th>Generic Drugs Name</th>
<th>Prescribing</th>
<th>Dispensing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simvastatin</td>
<td>331 (71%)</td>
<td>322 (97%)</td>
</tr>
<tr>
<td>Metoprolol succinate</td>
<td>266 (69%)</td>
<td>260 (98%)</td>
</tr>
<tr>
<td>Bendrofluazide</td>
<td>196 (66%)</td>
<td>186 (95%)</td>
</tr>
<tr>
<td>Cilazapril</td>
<td>175 (59%)</td>
<td>170 (97%)</td>
</tr>
<tr>
<td>Felodipine</td>
<td>123 (68%)</td>
<td>123 (100%)</td>
</tr>
<tr>
<td>Metformin</td>
<td>89 (60%)</td>
<td>87 (98%)</td>
</tr>
</tbody>
</table>

From Table 7.13 the extent to which patients who have good adherence based on prescribing also show good adherence based on dispensing is not evident. Therefore, for each drug, prescribing based adherence and dispensing based adherence was cross-tabulated (Table 7.14) to show PPV and NPV of non-adherence with prescribing (MPR_Pr <80%) for non-adherence...
on dispensing (MPR_Dp <80%), as well as to look at simple Pearson correlation of MPR_Pr and MPR_Dp.

Table 7.14: Association of individual MPR_Pr and MPR. (N = number of patient prescribed to as per Table 7.13, PPV = positive predictive value, NPV = negative predictive value, CI = confidence interval, R² = Pearson correlation between MPR_Pr and MPR_Dp.)

<table>
<thead>
<tr>
<th>Generic Drug Name</th>
<th>MPR_Pr</th>
<th>MPR</th>
<th>PPV</th>
<th>NPV</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt; 80%</td>
<td>≥80%</td>
<td>(95% CI)</td>
<td>(95% CI)</td>
<td></td>
</tr>
<tr>
<td>Simvastatin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 331)</td>
<td>80 (24%)</td>
<td>17 (5%)</td>
<td>0.825</td>
<td>0.782</td>
<td>0.3025</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.75-0.88)</td>
<td>(0.75-0.81)</td>
<td></td>
</tr>
<tr>
<td>Metoprolol succinate</td>
<td>&lt; 80%</td>
<td>54 (20%)</td>
<td>29 (11%)</td>
<td>0.651</td>
<td>0.787</td>
</tr>
<tr>
<td>(N = 266)</td>
<td></td>
<td></td>
<td>(0.57-0.73)</td>
<td>(0.75-0.82)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>≥80%</td>
<td>39 (15%)</td>
<td>144 (54%)</td>
<td>0.866</td>
<td>0.775</td>
</tr>
<tr>
<td>Bendrofluazide</td>
<td>&lt; 80%</td>
<td>58 (29%)</td>
<td>9 (5%)</td>
<td>0.789</td>
<td>0.798</td>
</tr>
<tr>
<td>(N = 196)</td>
<td></td>
<td></td>
<td>(0.78-0.92)</td>
<td>(0.73-0.80)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>≥80%</td>
<td>29 (15%)</td>
<td>100 (51%)</td>
<td>0.67-0.81</td>
<td></td>
</tr>
<tr>
<td>Cilazapril</td>
<td>&lt; 80%</td>
<td>53 (30%)</td>
<td>18 (11%)</td>
<td>0.746</td>
<td>0.798</td>
</tr>
<tr>
<td>(N = 175)</td>
<td></td>
<td></td>
<td>(0.67-0.81)</td>
<td>(0.74-0.84)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>≥80%</td>
<td>21 (12%)</td>
<td>83 (47%)</td>
<td>0.692</td>
<td>0.857</td>
</tr>
<tr>
<td>Felodipine</td>
<td>&lt; 80%</td>
<td>27 (22%)</td>
<td>12 (10%)</td>
<td>0.692</td>
<td>0.857</td>
</tr>
<tr>
<td>(N = 123)</td>
<td></td>
<td></td>
<td>(0.57-0.79)</td>
<td>(0.80-0.90)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>≥80%</td>
<td>12 (10%)</td>
<td>72 (58%)</td>
<td>0.692</td>
<td>0.857</td>
</tr>
<tr>
<td>Metformin</td>
<td>&lt; 80%</td>
<td>27 (30%)</td>
<td>9 (10%)</td>
<td>0.750</td>
<td>0.755</td>
</tr>
<tr>
<td>(N = 89)</td>
<td></td>
<td></td>
<td>(0.63-0.84)</td>
<td>(0.67-0.82)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>≥80%</td>
<td>13 (15%)</td>
<td>40 (45%)</td>
<td>0.750</td>
<td>0.755</td>
</tr>
<tr>
<td>Overall</td>
<td>&lt; 80%</td>
<td>241 (37%)</td>
<td>55 (9%)</td>
<td>0.814</td>
<td>0.763</td>
</tr>
<tr>
<td>(N = 646)</td>
<td></td>
<td></td>
<td>(0.78-0.85)</td>
<td>(0.73-0.79)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>≥80%</td>
<td>83 (13%)</td>
<td>267 (41%)</td>
<td>0.814</td>
<td>0.763</td>
</tr>
</tbody>
</table>

Patients in the cohort often possess multiple medications from among the six drugs (with frequencies shown in Table 7.15). The last row in Table 7.14 represents overall adherence rates to prescribed and dispensed medication after taking the multiple usage of Table 7.15 into account, where a patient is considered non-adherent overall if MPR<80% for any of the six drugs for which the patient had a prescription extending into or during the 15-month period. By this measure, only 50% of patients (322 of 646) were adherent to all their prescribed medications (of the six we track) based on dispensing. Overall PPV and NPV indicate that if non-adherence is indicated by the MPR_Pr, a patient is 81% likely to show non-adherence by MPR_Dp; and if a patient shows high adherence by MPR_Pr, the patient is 76% likely to show high adherence by MPR_Dp.
Table 7.15: Frequencies with which patients are prescribed and dispensed multiple of the six selected drugs during the evaluation period, or with supply extending into the evaluation period.

<table>
<thead>
<tr>
<th>Number of Drugs</th>
<th>Prescribed</th>
<th>Dispensed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>286</td>
<td>249</td>
</tr>
<tr>
<td>2</td>
<td>224</td>
<td>209</td>
</tr>
<tr>
<td>3</td>
<td>96</td>
<td>106</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>56</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

In addition to PPV and NPV, another way to understand the relationship of MPR_Pr with MPR_Dp is to develop a calibration curve. For this analysis, ‘bins’ at 5% intervals were created for overall MPR_Pr and the average overall MPR_Dp was determined for all patients who had an overall MPR_Pr that fell within that MPR_Pr range (i.e., the ‘bin’ of interest). The resulting calibration curve is shown in Figure 7.5. The cumulative frequency of prescriptions by MPR_Pr area is also plotted on Figure 7.5, highlighting that the data is relatively sparse until MPR_Pr reaches around 50%. Note that two data points are omitted from the calibration curve as outliers (MPR_Pr=9.27%, MPR_Dp=100% and MPR_Pr=10.68%, MPR_Dp=40.40%); with these included the MPR_Dp at MPR_Pr=10% would be 48.98%.

Figure 7.5: Calibration curve for overall MPR_Dp vs. MPR_Pr with best linear estimate (omitting two outliers in the MPR_Pr = 10% bin); and cumulative frequency of overall MPR_Pr (dashed line)
7.7.7 First Dispensing of a Prescription

Lacking unique identifiers for prescriptions, it is not possible to directly identify which prescription was being filled by a specific dispensing record. However, it was deemed that a dispensing record within a seven-day period after prescribing of the same generic drug to the same patient to be consistent with adherent behaviour with respect to the first dispensing of a given prescription. Table 7.16 shows the distribution of prescription fill rates for the selected drugs.

Table 7.16: Prescriptions with associated first dispensing within 7 days after being prescribed

<table>
<thead>
<tr>
<th>Generic Drug Name</th>
<th>Number of Prescriptions</th>
<th>Dispensing within Seven Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simvastatin</td>
<td>1159</td>
<td>1072 (92%)</td>
</tr>
<tr>
<td>Metoprolol succinate</td>
<td>949</td>
<td>917 (97%)</td>
</tr>
<tr>
<td>Bendrofluazide</td>
<td>624</td>
<td>556 (89%)</td>
</tr>
<tr>
<td>Cilazapril</td>
<td>566</td>
<td>526 (93%)</td>
</tr>
<tr>
<td>Felodipine</td>
<td>426</td>
<td>403 (95%)</td>
</tr>
<tr>
<td>Metformin</td>
<td>319</td>
<td>286 (90%)</td>
</tr>
</tbody>
</table>

The high rate of presence of a dispensing record within seven days of prescribing (93% on average) indicates that patients generally receive their first dispensing from a prescription in a timely manner, and that the non-adherence detectable in MPR_Pr, but not MPR_Dp, is usually related to failure to return for refills.

7.7.8 Relationship to other Potential Adherence Predictors

While MPR_Pr shows substantial predictive value on MPR, there is a question of how much information it provides beyond other indicators available to a general medical practice. Demographic characteristics of the patients prescribed the selected medications, as indicated in the practice’s records in the PMS, are summarised in Table 7.17. Maori and Pacific Island ethnicity, and lowest socioeconomic quintile (based on census data and patient address), relate to the PHO funding formula such that the practice would be motivated to code these values.
Table 7.17: Characteristics of the patients on one or more of the six long-term medications (N = 646)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Number of Patients (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt; 30</td>
<td>4 (1%)</td>
</tr>
<tr>
<td>30-44</td>
<td>41 (6%)</td>
</tr>
<tr>
<td>45-59</td>
<td>169 (26%)</td>
</tr>
<tr>
<td>60-74</td>
<td>234 (36%)</td>
</tr>
<tr>
<td>75+</td>
<td>198 (31%)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>397 (61%)</td>
</tr>
<tr>
<td>Male</td>
<td>249 (39%)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Maori</td>
<td>47 (7%)</td>
</tr>
<tr>
<td>Pacific</td>
<td>12 (2%)</td>
</tr>
<tr>
<td>European</td>
<td>561 (87%)</td>
</tr>
<tr>
<td>Asian</td>
<td>11 (2%)</td>
</tr>
<tr>
<td>Other</td>
<td>15 (2%)</td>
</tr>
<tr>
<td><strong>Most deprived quintile</strong></td>
<td>121 (19%)</td>
</tr>
</tbody>
</table>

A linear regression model of patient MPR_Dp was developed using the above factors (coding age in years and others as binary indicators), along with binary factors based on MPR>0 for each of the six drugs. Regression was on arcsine of square-root of MPR_Dp (a commonly used transformation for proportional data). The model resulted in an adjusted $R^2$ of 0.161 and a standard error of 0.301. Gender, five of the six drugs (excepting metformin) and European ethnicity were statistically significant coefficients.

Recomputing the regression with MPR_Pr in addition to the other variables resulted in an adjusted $R^2$ of 0.434 and a standard error of 0.248, with, aside from MPR_Pr, only gender and four of the drugs (simvastatin falling out) having significant coefficients. The influence of MPR_Pr on transformed MPR_Dp (coefficient=0.804, 95% CI 0.71-0.89, $p=8.5 \times 10^{-56}$) was approximately 10 times greater than the most significant drug factor (bendrofluzide, coefficient = 0.0833, 95% CI 0.039-0.128, $p=2.8 \times 10^{-4}$) and nearly 18 times greater than the significant demographic factor (female gender, coefficient=0.045, 95% CI 0.003-0.086, $p=0.036$), indicating MPR_Pr as the dominant factor in predicting MPR_Dp.

**7.7.9 Interpretation**

This study assessed the correspondence of general medical practice prescribing data to community medication dispensing records for six government funded long-term medications frequently used in chronic condition management. The analysis has permitted to see that there is very substantial, but by no means complete, information about dispensing conveyed in the
prescribing records for these long-term medications. Medication possession level as estimated from prescriptions, MPR_Pr, was observed to provide 81% PPV and 76% NPV for dispensing based non-adherence (MPR_Dp <80%). The MPR_Pr is observed to be a far more powerful predictor of MPR_Dp than demographics or socioeconomic status, although additional adherence prediction information is provided by the specific drugs prescribed. The results also indicate, as expected, that adherence problems in the community are very commonplace, occurring in about half the cases with respect to the six long-term medications analysed during a 15-month interval. This high prevalence of non-adherence provides a fertile ground for useful interventions.

7.8 Improving GP awareness of antidepressant adherence issues

The various studies discussed thus far have focussed primarily on antihypertensive medication adherence. However, as briefly discussed in Section 5.5.2, ChronoMedIt’s knowledge base is easily extensible to other clinical domains as well, provided that the nature of queries that need to be executed fall into one of the four criteria classes. The purpose of this study was to demonstrate the use of ChronoMedIt within a different domain to raise GP awareness of antidepressant adherence issues. The proposition came about as a psychiatrist became interested in the work I have been doing in the antihypertensive space.

Depression is among the most treatable of mental disorders – between 80-90% of people eventually respond well to treatment and return to their normal lives [276], yet a major issue in the management of depression is poor adherence to antidepressant therapy [277]. Treatment guidelines provide guidance on optimal choice of medicine [278] and guidelines such as the ones by NICE in the UK [279] and the NZ Guidelines Group [280] suggest a minimum of six months of continuous antidepressant therapy for the first episode of depression and longer for subsequent episodes. However, research indicates that 44% of the patients stop taking medication by the third month of therapy [281].

Most people with depression are treated in a general practice setting, either by a GP alone, or (for more serious depression) in partnership with specialist mental health services [282]. This provides an opportunity to use general practice EMR data to detect patients who show poor adherence to antidepressant therapy (and hence on suboptimal therapy), a natural extension to my work around antihypertensives.
7.8.1 Analysis Protocol

Unlike a chronic condition such as diabetes, many patients with depression can be successfully withdrawn from treatment after an initial treatment period and therefore do not need to be on antidepressants for the rest of their lives. Therefore, a six month period from 1-Jul-06 to 31-Dec-06 was used as the evaluation period and only patients having an antidepressant-free period of six months prior to the start of the evaluation period were included, i.e., no antidepressant prescription present during the period from 01-Jan-06 to 30-Jun-06. A gap-based adherence measure was used as an indication of the time of continuous therapy and a maximum ‘permissible gap’ of 15 days was allowed where patients could go without a dose and not anticipate reduced or suboptimal outcomes [283]; patients who have gaps (also referred to as ‘lapses’ previously) in treatment exceeding this duration are deemed non-adherent. Using this protocol and the same prescribing-dispensing dataset mentioned in Section 7.7.3, prescribing of antidepressant medicines is examined together with their community pharmacy dispensing. Similar to the previous studies, only funded patients enrolled at the practices were included.

7.8.2 Medication Lapses in Antidepressant Therapy

Prescribing data was first analysed to identify adherence issues in antidepressant therapy, as indicated in the practice’s EMR. 153 patients satisfied the inclusion criteria and were therefore inferred to have started their antidepressant therapy during the evaluation period. Out of these patients, 59 (39%) patients were identified as being non-adherent.

Using ChronoMedIt’s prescription visualisation tool, commonly occurring antidepressant prescribing patterns for non-adherent patients are shown in Figure 7.6.
Figure 7.6: Commonly occurring patterns related to non-adherence. The tooltip in the top figure shows further details related to the selected prescription.

The top timeline in Figure 7.6 shows a case where there was only a single 90-day prescription. The second case is consistent with the GP deciding to start the patient on a short, 30-day course (perhaps to check on efficacy and whether the patient had any adverse reactions) and then prescribing a normal 90-day script upon returning for the next prescription. However, this patient has not had a full six-months of antidepressant therapy, and hence the guideline has not been complied with. The last case is a typical scenario where the patient appears to have failed to return for the second prescription on time, possibly as a result of not consuming the medicines from the previous prescription as directed by the GP. These three cases represent three broad categories of non-adherence, which were identified as (1) failure to return for re-prescription, (2) failure to maintain adherence despite initial attempts (note that there is only a single re-prescription for 90-days with respect to the second case in Figure 7.6, but this could have even been three 30-day re-prescriptions, but still having the same effect), and (3) failure to return for re-prescription in a timely manner, respectively. Via visual inspection of the individual cases, all other cases of non-adherence were concluded to be a variation or a combination of these three categories.
Figure 7.7 shows several cases related to adherent patients. The first case satisfies the minimum guideline requirements while the other two patients have been on antidepressants for a longer duration. The second patient in Figure 7.7 and the second patient in Figure 7.6 have some similarity as the patient was started on a shorter 30-day duration script and then moved onto a standard 90-day script, and it is possible that the prescribing pattern seen in the second patient in Figure 7.7 was the GP’s intention for the second patient in Figure 7.6 as well.

![Figure 7.7: Several prescribing patterns for adherent patients](image)

### 7.8.3 Prescribing vs. Dispensing Based Adherence

Similar to the previous study, the patients satisfying the inclusion criteria were also matched to their dispensing via NHIs to determine PPV and NPV of non-adherence with prescribing for non-adherence on dispensing (Table 7.18).

<table>
<thead>
<tr>
<th>Prescribing</th>
<th>Dispensing</th>
<th>PPV (95% CI)</th>
<th>NPV (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;15 day lapse</td>
<td>58 (38%)</td>
<td>1 (1%)</td>
<td>0.98</td>
</tr>
<tr>
<td>No lapse</td>
<td>46 (30%)</td>
<td>48 (31%)</td>
<td>(0.92-0.99)</td>
</tr>
</tbody>
</table>
Using age, gender, being in the lowest socioeconomic quintile and ethnicity as independent variables, a multivariate model was then constructed to determine what factors were associated with adherence to antidepressant dispensing for the cohort of 153 patients. Out of these, 21 patients had only prescribing records (i.e., no dispensing), and a >15 day lapse in dispensing was considered for them. Characteristics of the included patients are shown in Table 7.19.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Number of Patients (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>&lt; 30</td>
<td>23 (15%)</td>
</tr>
<tr>
<td>30-44</td>
<td>36 (24%)</td>
</tr>
<tr>
<td>45-59</td>
<td>45 (29%)</td>
</tr>
<tr>
<td>60-74</td>
<td>31 (20%)</td>
</tr>
<tr>
<td>75+</td>
<td>18 (12%)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>111 (73%)</td>
</tr>
<tr>
<td>Male</td>
<td>42 (27%)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>Maori</td>
<td>8 (5%)</td>
</tr>
<tr>
<td>Pacific</td>
<td>2 (1%)</td>
</tr>
<tr>
<td>European</td>
<td>138 (90%)</td>
</tr>
<tr>
<td>Asian</td>
<td>1 (1%)</td>
</tr>
<tr>
<td>Other</td>
<td>4 (3%)</td>
</tr>
<tr>
<td>Most deprived quintile</td>
<td>30 (20%)</td>
</tr>
</tbody>
</table>

The Adjusted R² value of this model was -0.0098; none of the independent variables were statistically significantly associated with dispensing based adherence rates.

This model was then extended by including prescribing based adherence rates for each patient. This model resulted in an F-test value of 6.2 (p < 0.001) indicating that the overall model is significant. However, when the individual variables were considered, only prescribing based adherence was associated with a statistically significant p-value (coefficient=0.44, 95% CI 0.31-0.58, p = 1.2 x 10⁻⁹). Adjusted R² in this model was 0.2143 which indicates that the inclusion of prescribing based adherence has improved the model (but by no means fully explains all patient behaviour with respect to dispensing).

### 7.8.4 Implications

This study analysed antidepressant adherence and non-adherence rates using general practice prescribing data and pharmacy dispensing data. A main goal of this study was to start developing a process to inform GPs to become more aware of non-adherence in the community.
and demonstrate some commonly occurring types of prescribing patterns. Three broad patterns related to non-adherence were identified and the use of the prescription visualisation tool was demonstrated as a potential tool to support clinicians in seeing the nature of adherence problems as indicated in the data (which may be difficult to discern in a tabular list of prescribing dates). The results indicate that poor adherence rates were 39% and 68% based on prescribing and dispensing, respectively indicating that EMR prescribing data can be used to assist GPs to have more informed conversations with patients.

7.9 A feasibility study of nurse-led adherence promotion

Our research has confirmed that there is a large cohort of high-needs patients in NZ with demonstrated risk related to adherence identifiable through general practice PMS records. The results presented in Sections 7.2 – 7.4 indicate that antihypertensive medication adherence is poor and usually around 40%. As discussed in this thesis, being able to identify such patients presents an opportunity to intervene on them as appropriate.

In NZ, the obvious foundation for systematic follow-up for adherence promotion called for by Fahey et al. [284] is through exploitation of the information in the general practice PMS. The ChronoMedIt framework implementation has been verified (see Chapter 6) and there has been interest from our collaborating clinicians to carry out a feasibility study (which is also a field validation exercise of some of ChronoMedIt’s audit capabilities). As a result, a feasibility trial was designed to evaluate the potential to improve BP through improved adherence to medication in patients (identified using ChronoMedIt) indicated in the PMS data using a systematic nurse-led telephone outreach operating within the general practice environment. The intervention and methodology has recently been funded by the NZ Health Research Council (HRC) to the tune of $150,000 and it is currently being trialled and refined within a ‘feasibility study’ approved under the HRC protocol 09/136R (Adherence Innovations in Medication Use for Health Improvement, AIM-HI). The main aims of this feasibility trial are:

1. To assess the feasibility of PMS-based monitoring of prescribing to inform telephone outreach as an intervention to improve timely receipt of prescriptions for BP lowering medication.

2. To pilot this intervention over a six-month period as the basis for a future cluster RCT that will demonstrate significant clinical outcomes with respect to BP and renal function.

The study has been designed as a two arm feasibility trial and following are some related details of the study:
Study population and setting: Patients in two Pacific Island led general practices: Practice A, West Fono Health in Waitemata DHB (WDHB); Practice B, South Seas in Counties Manukau DHB (CMDHB). PMS data was extracted and measured for all enrolled patients over 20 years of age who had had one or more prescriptions for BP lowering medication from the practice between 18 and 12 months preceding the start of the intervention. We received support from collaborating clinicians from West Fono and South Seas Healthcare to carry out this intervention as a pilot study.

Intervention: In Practice A, PMS-data was used to identify patients with poor history of timely re-prescription. This practice has undertaken telephone ‘outreach’ calls for the identified patients. Patients of Practice A with an MPR for antihypertensives of less than 80% for the previous year and who lacked confirmed current controlled BP were the target for this outreach. The outreach programme was applied by nursing staff with high cultural competency apropos to the Samoan, Tongan, Niuean, Tuvaluan and Cook Island populations. Outreach calls encouraged practice attendance for assessment and medication prescriptions as well as addressing potential barriers to adherence using behaviour modification techniques. Contact methods escalated from land lines, to mobile phone, to home visiting if required. The nurses are currently working through the list of intervention patients to a fixed time budget, keeping notes on success rates in reaching patients and issues encountered. Practice B is continuing with ‘usual care’.

Outcome measures:
1. Timely re-prescribing as improvement in MPR.
2. Medications taken (based on pill counts identified at a sample of patient homes)
3. Systolic and diastolic BP (as recorded in PMS)
4. Five-year cardiovascular risk (as computed on a sample of patients)
5. Renal function: serum creatinine, eGFR and uric acid as per lab results transmitted to the PMS by HealthLink

At the end of the study, 40 outreach patients from the intervention practice will be randomly selected and invited for home visits that will assess pill counts and determine the alignment of available medication to adherence with the prescriptions during the first and last quarters of the year.
Qualitative Arm: Focus group feedback will be gathered from both consenting patients and the outreach nurses to refine the appropriateness and effectiveness of the outreach, including information needs of the outreach nurses, content of the outreach call, and the nature of follow-up actions. The patient focus groups will follow on from data already collected based on semi-structured face-to-face interviews with 20 participants. A further four groups of five outreach patients will be recruited for focus group discussions on both their experience of adherence issues and the appropriateness and effectiveness of the intervention, one in each quarter of the feasibility study. The major outcome variables to be collected during this study are reasons for not collecting medication on time or for not taking medication according to the prescription. The nurses will convene a quarterly focus group to share experiences and refine their outreach methods. Transcripts will be analysed using the general inductive approach [285] to identify themes using grounded theory to inform refinement of the intervention.

Confidentiality: The system of electronic medication monitoring can be run such that non-practice researchers (including myself) are blind to the patient identity and a practice identifier enables practice staff to identify those who need telephone outreach. The outcome measures report percentages of patients within the practice population or sub-population so individuals are not identifiable.

The feasibility study intervention is currently being trialled only with two practices, one as intervention and another as control, but two other Waikato area general practices are also taking up the intervention to test our ability to port the methods to settings outside their initial development and to expand the qualitative data sources.

The findings based on preliminary analyses include that:

1. Reasons for non-adherence include family and social commitments, work schedules that interfere with medication administration and transport issues, as well as forgetfulness;
2. Patients with good adherence express understanding of the importance of their medication, good communication and trust with their GP, and effort at time management, as well as having supportive family and, frequently, a previous cardiovascular event;
3. Patients vary in what constitutes acceptable and feasible means of being contacted, including variability in cell phone possession and use (notably, whether it is kept on person and charged), variability in impact of postal reminders and variability in interest
in home visit. However, telephone reminders on the order of a week in advance of a visit are very broadly accepted and valued;

4. Telephone contact of patients with MPR <80% is largely successful, requiring around two calls to make contact.

Co-investigators in the AIM-HI study have explored influences on adherence to taking long-term medications among Samoan patients [30] within one of the practices in the study. I provided the lead-investigator with a list of 20 Samoan patients (10 patients with MPR ≥80% and 10 with MPR <80%) from the practice who were then interviewed about their views on adherence or non-adherence to taking antihypertensive medications. One-to-one semi-structured interviews using open-ended questions were undertaken in Samoan and English, recorded, transcribed and translated into English. Transcriptions were examined by two researchers to identify themes. The main themes that have been identified for those with low adherence are ‘lack of transport’, ‘family commitments’, ‘forgetfulness’, ‘church activities’, ‘feeling well’ and ‘priorities’ while the themes identified for those with high adherence rates are ‘prioritising health’, ‘previous event’, ‘time management’, ‘supportive family members’ and ‘relationship with GP (language and trust)’. A theme common to both was ‘coping with the stress of multiple co-morbidities’.

Other than intending to analyse the data from the AIM-HI feasibility trial once it is over, there are also plans to extend the protocol to a much larger RCT – some details of this proposed RCT are discussed in Section 8.4.
Chapter 8

Discussion

The main purpose of this chapter is to discuss the significance of the work presented in this thesis, discuss other related work, address current limitations that have been identified and discuss intended future directions.

8.1 Significance

A main contribution of this thesis is the development of ChronoMedIt, a novel computational framework that can be used to identify patients on long-term medication who are on suboptimal therapy. The generic framework has been designed to be flexible, easily extensible and grounded in an underlying well-defined Criteria Model. There has been considerable emphasis on the flexibility in the parameters of the various queries to suit a range of guidelines and specific circumstances under which clinical teams may wish to conduct clinical audit, as well as in consideration of how much is yet to be learned about interventions to improve adherence. The XML document format that has been created for queries provides a transparent and standards-based interface to the architecture (along with the use of Protégé-OWL to manage relevant concept taxonomies for drugs and problem classifications). A prime focus of the research has been on providing practice-specific and/or patient-specific information to make it feasible for practice staff to assess the quality of their management of patients on long-term medications over time. This leads to two important distinctions for quality improvement efforts. Firstly, the auditing can be used to support assessment of a quality improvement program at a practice level where performance within the practice can be assessed (e.g., what
percentage of the practice population with hypertension had at least three consecutively high BPs during the past year?); this could primarily be to determine how well the practice is doing for internal quality audit purposes and to provide feedback to clinicians. Secondly, the detailed patient level information provided in the audit reports can be used to examine the specific cases exhibiting suboptimal management and to intervene on these patients and also provide patient-specific feedback/education.

The nature of chronic illness requires reasoning on time intervals, which are highly implicit in PMS databases that store time-stamped events. Thus, capturing the clinical requirements necessary to characterise good chronic disease management leads to highly-complex and potentially error-prone queries. As discussed in the previous chapters, particularly in Chapters 3-5, the implementation of the framework presented in this thesis carefully takes various temporal considerations into account when identifying patients on suboptimal therapy, and as such, can be used to cater for domain-specific querying requirements of quality improvement for general practice activities. Accuracy of the various query results has been verified using independent implementations, and random testing has been performed on 10,000 simulated test patients to ensure software reliability. The framework has already been used in several ‘real world’ situations, for instance, the validation study discussed in Chapter 3 and the feasibility study discussed in Section 7.9, demonstrating the validity/suitability of the developed framework to be used in realistic clinical environments.

ChronoMedIt uses an ontology based representation for drugs and classifications, which not only makes it easier to create, visualise and navigate the often complex and hierarchical knowledge bases, but also provides the ability to cluster many transaction-level concepts (e.g., classifying a patient using Read Code G2.00 and writing a prescription for cilazapril) into more domain level (possibly hierarchical) concepts. Examples include cilazapril drug being included under the ACEi/ARB drug class, which in turn belongs to antihypertensive medications; and the Read Code G2.00 being associated with the domain level concept ‘hypertension’. Having this type of knowledge level support facilitates domain level querying, for instance, hypertension can be specified as a classification directly when querying instead of having to specify the individual transaction-level codes that constitute hypertension. The ontology based approach also improves manageability and reusability of the knowledge bases as well as making it easier to extend the domain of application of the framework; for example, expanding the relevant classification codes to other schemes (such as SNOMED, ICD-9 and ICD-10) by editing the knowledge bases in Protégé-OWL.
In addition to the computational engine, ChronoMedIt also consists of a novel graphical tool that can be used to visualise prescribing patterns using high level concepts across all the prescribed drug classes (such as antihypertensives, ACEis, ARBs or antidepressants). The tool can also be used to visualise various measurements, such as BPs, making it potentially easier to understand how a patient’s clinical outcomes have changed with therapy over time.

The different studies based on production EMR data discussed in Chapter 7 indicate that:

1. a significant portion (59% and 63% respectively for the two practices concerned) of patients with hypertension have >30 day lapses in their antihypertensive medication; and over a third of people with hypertension have not had a BP measurement for >180 days;

2. at least 56% of patients with hypertension and diabetes showed poor adherence to ACEi/ARB therapy (MPR <80%), although as a result, these patients were more likely to have uncontrolled BP than adherent patients (odds ratio = 4.0, \( p = 0.002 \) and odds ratio = 2.5, \( p = 0.034 \) for the two practices);

3. adherence to antihypertensive therapy is correlated to having controlled BP with non-adherent patients being more likely to have uncontrolled BP than adherent patients (odds ratio = 2.4, \( p = 0.001 \) and odds ratio = 1.7, \( p = 0.03 \) for the two practices); mean reductions in systolic BPs were observed to be 19.31 mmHg and 16.39 mmHg respectively for the two practices for being adherent from 0% to 100%;

4. interval based measures, such as MPR, are more stable measures than single, point-in-time measures in identifying patients with poor BP control;

5. satisfying a single, point-in-time measure may not necessarily be an indication of optimal management of BP and other measures need to be considered, especially if quality indicators are associated with financial incentives;

6. analysing prescribing data has much merit and can provide 81% PPV and 76% NPV for dispensing based non-adherence; and

7. 39% of the patients starting antidepressants were found to be non-adherent and it was shown that using the prescription-visualisation tool may provide an opportunity for clinicians to have more informed conversations with patients.

As such, ChronoMedIt has already been used in several different clinical contexts and shows potential to function as a computational quality audit reporting framework that can be used to
formulate clinically important queries. Two broad forms of reporting for improvement of chronic disease management are evident from this work:

1. If an agreed set of audit improvement criteria can be established, this form of reporting provides an opportunity to provide feedback to GPs on aggregate patient data to improve the management of their patients with chronic conditions. This also provides a ground to compare general practices (as done in some of the studies in Chapter 7 with two practices) which in turn can lead to patient-provider education type interventions for the poorly performing practices if needed.

2. Awareness of individual cases – addressing this dimension of reporting can take three broad directions; (i) patient related issues – identification of those patients who, at a particular moment in time, are on suboptimal management (for instance, out of supply of an indicated medication or no BP measurement after a high BP). In the first instance, the action is to treat the non-adherence to clinical advice (e.g., poor medication adherence, not scheduling an appointment to see the GP in a month) as inadvertent and recall the patient and/or simply take appropriate action, such as prescribe as indicated or measure BP, at the next opportunity. This includes not only patients on suboptimal management, but also those whose circumstances have changed (e.g., due to development of a co-morbidity) and thus require additions to previous therapy; (ii) GP-patient interaction related issues – with respect to individual patients with poor management profiles, there is an opportunity for improved GP-patient communication since at some point it becomes logical to look at the lack of concordance between doctor and patient. Poor management over an extended time period (e.g., repeated lapses in medication supply) indicate the need for enhanced communication between GP and patient, possibly the clinician needs to engage the patient more in a joint ‘problem-solving’ approach in relation to underlying barriers; (iii) GP related issues – this form of reporting also provides an opportunity to critique GPs on their adherence to established guidelines, for example, the JNC7 hypertension guideline [6] recommends ACEi/ARB medication as compellingly indicated therapy for patients with comorbid hypertension and chronic kidney disease. If the quality improvement reports indicate that patients are not on such indicated therapy, it is possible that it was due to the GP not prescribing these medications. As such, investigating GP adherence to guidelines is another possibility.

As a broader research opportunity, this research has shown that GP EMR data is a promising resource for gaining a more detailed understanding of the factors that predispose patients to
non-adherence risk and thence for development of targeted intervention strategies for specific clusters of patients (e.g., those who are persistently inadvertently non-adherent due to specific lifestyle issues versus those who are intentionally non-adherent due to disagreeing with the doctor’s recommendations or simply due to cost).

8.2 Related Work

8.2.1 Frameworks and systems

Statistical reporting and retrospective auditing of general practice medicine is not uncommon and there have been many past studies, for example, where patient non-adherence to medication has been discussed [183, 286, 287]. However, often, the related patient data was collected specifically for the purposes of carrying out a quality audit; for instance, handing out electronic monitoring devices to patients and determining medication adherence for patients who returned the devices [288], and developing manual data abstraction protocols and training nurses to carry out the data abstraction process [183]. What is distinct in the current approach is that it offers a valuable, flexible architecture that is not tied to a specific query and can be used to formulate a range of criteria within the scope of the criteria classes. The architecture can carry out audits directly from the EMR and present not only practice-specific details, but patient-specific details as well. Most PMS systems have ‘query-builder’ reporting functionalities for day-to-day reporting; however, the full consideration of the temporal issues for concepts such as lapse and MPR is non-trivial (as discussed in Chapters 3-5), as is the knowledge-base requirements to produce domain level concepts (e.g., ‘antihypertensives’ and ‘diabetes’) from records that code specific drugs and diagnosis codes. Despite the computational challenges to achieving precise measures of chronic disease management activities, we have entered an era where quality auditing (and, in fact, ‘pay for performance’) are operational realities in some healthcare systems.

The QOF in the UK [196] is part of the General Medical Services contract which encourages GPs to use evidence-based interventions, particularly in the management of chronic diseases (such as diabetes and hypertension). The linking of the QOF directly to GP payment and its clinical outcomes has had mixed reviews with [221, 289] reporting improvement in certain conditions while [290-292] have viewed the whole scheme with more skepticism. Irrespective of my opinion of the UK based implementation, it is important to give individual jurisdictions the ability to create their own regional/local level indicators – for example, the Maori and Pacific Islander populations in NZ have inherently higher CVD risk, all other factors being equal, than
their European counterparts [293]; thus, defining a ‘controlled BP’ as 130/80 mmHg will be more appropriate than the 145/85 mmHg level as used in the DM12 indicator of the QOF. It is interesting to note that an indicator with one of the most QOF point allocations is for achieving controlled BP (at the level of 150/90 for patients diagnosed with hypertension, and with additional points for tighter management of patients with diabetes and with chronic kidney disease). Since MPR is associated with improved odds of successful BP control (at the more stringent targets recommended in JNC7), and also shown to be a more stable ‘interval-based’ measure than a ‘point-in-time’ based measure as used in the QOF, there may be an opportunity to improve QOF performance using an analysis framework similar to the one presented herein.

The developed criteria can be used to identify patients who need a follow-up (e.g., ‘Patients with hypertension who have a high BP measurement followed by a lapse of 120 days’). With respect to the QOF, Section 7.6 indicated how patients with hypertension or diabetes may satisfy two of the important QOF criteria while not having optimal BP control due to using single point-in-time measures. Similarly, QOF CHD6 indicator is “…patients with coronary heart disease [CHD] in whom the last blood pressure reading (measured in the previous 15 months) is 150/90 or less”, but if a patient with CHD had the last BP measurement (which was also ‘controlled’) almost 15 months ago, and then no BP measurement at all, such a patient would satisfy CHD6, although the patient’s BP may no longer be controlled. Therefore, to improve the patient follow-up process it would be beneficial to also use a criterion that is related to timely BP measurement in conjunction with CHD6.

Antihypertensive therapy has been the subject of considerable decision support systems research, notably the PRODIGY [15] and ATHENA [14] systems. An underlying factor common to ATHENA, PRODIGY and most other decision support tools is that they all attempt to computerise relevant CPGs and let the GP prescribe according to the system recommendations, hence using effectively a knowledge engineering approach. Although almost all the systems have the capacity to let the GPs override the system recommendations and have critiquing modes where the system interrupts only when critical alerts need to be made, ChronoMedIt takes the position that auditing general practice with specific criteria is (or possibly even more) beneficial for the management of patients with chronic illness. This position can be supported by the complexities and difficulties involved with chronic illness where various temporal reasoning mechanisms are required [18] and other associated problems reported during PRODIGY phase II implementation [15]. Therefore, the present proposition differs chiefly in the emphasis on evaluating deficiencies in patient management over a time period/interval (i.e.,
occurring *anytime* during that period of interest) than looking at deficiencies at the time of consultation; and with application to quality improvement over a cohort of patients rather than as interactive decision support at the point of care. In fact, the work presented in this thesis most closely resembles IDAN/KNAVE-II framework [143, 145] and the VISITORS system [146, 148] for temporal abstraction and visualisation on clinical data than the automated-guideline approach as represented by PRODIGY and ATHENA. However, this work takes a much narrower, focussed/targeted approach built around specific needs for chronic disease management than a generic approach which has allowed the majority of the work to be done (after sufficient conditioning of the data in our architecture) through a set of domain specific SQL queries – for example, using the IDAN/KNAVE-II framework one can abstract/visualise higher level concepts (e.g., LOW, HIGH) according to a given context using raw, time-stamped data (see Figure 2.6); however, developing such generic abstractions using constraints specified within a temporal ontology has not been part of this work. Rather, this work explored in greater detail the specific temporal issues around cohorts of individuals over fixed ‘evaluation periods’ while considering the interplay of therapy and indications. On the other hand, the HyperCritic system (see Section 2.2.4) uses a computerised representation of a CPG as its underlying model and critiques GP actions as relevant to hypertension management using only data in the EMR. HyperCritic focusses on critiquing individual patients at the point of care, effectively providing decision support via its critiques. When evaluating HyperCritic, the panel of reviewers assessed the critiques generated by the system and compared them against comments made by physicians (i.e., a system vs. physician type evaluation). This evaluation method has some resemblance to the 20/20 study discussed in Section 3.4 which had a relatively similar setting and assessed to some degree a GP’s ability to identify suboptimality in therapy given a patient; however, the primary focus of the 20/20 study was to validate the identified criteria and to ensure ChronoMedIt could accurately identify patients meeting a given criteria. Other than similarities on evaluation, ChronoMedIt also uses only data in the EMR, but has placed emphasis on critiquing cohorts of patients using an agreed set of indicators, thus providing the opportunity to benchmark practices on these indicators.

The Medical Database Adaptor (MEIDA) [138] is a framework that is focussed on facilitating the use and reuse of decision support systems and knowledge bases with heterogeneous clinical databases by overcoming the heterogeneity of the clinical databases. The motivation behind this work was that often the terms that one decision support system uses to describe patient information is not recognised (with the same interpretation) by another decision support system or a local clinical database. As a possible solution, MEIDA proposes a virtual schema for patient
data which enables interoperability between a decision support system knowledge base and a given EMR structure. MEIDA uses standardised terminologies (such as ICD-9) to map between local database terms and the virtual EMR, and the mapping process used in this thesis work is somewhat similar, but more application specific – this thesis work used a ChronoMedIt specific schema to describe patient data and then mapped details from the PMS data to the database during the data extraction stage. MEIDA provides support to automate the local database mapping process and as such if a similar technique could be developed to eliminate the data extraction stage and work directly with the PMS’s database, it will be beneficial. However, PMS vendors often have proprietary databases and providing such close integration may be a challenge.

Jin et al. [162, 163] propose a novel data mining oriented approach using unexpected temporal association rules (UTARs) where adverse drug reactions that occur within a given time period (somewhat similar to the evaluation period) can be detected directly from healthcare administrative databases while Fails et al. [294] have proposed PatternFinder, a novel querying tool that can be used to specify and discover temporal patterns, and also visualise the results as relevant to adverse reactions [295]. These are approaches where the adverse drug reactions are ‘discovered’ automatically, whereas the ChronoMedIt framework developed within the scope of this thesis gives a careful quantification of a known issue as compared to discovering a new issue. Hence, there is a parallel of this work to techniques used within UTARs and PatternFinder in that they have specialised methods for adverse drug reaction detection, while ChronoMedIt has specialised methods for adherence and other quality audit assessment.

There has also been an attempt “to develop an explicit tool precisely for measuring guideline adherence” in the context of adherence to the JNC7 guideline where the researchers developed 22 explicit criteria in four domains of care [183]. This tool, however, is focused on an overall scoring of adherence of the care process to the guideline, including quality of documentation, as compared to ChronoMedIt’s focus on quantified targets around specific areas of successful chronic disease management. With somewhat similar intentions, Advani et al. [296, 297] have focussed on developing an approach (with related algorithms to calculate adherence scores) to evaluate and consistently score clinician adherence to CPGs using the intentions of guideline authors. They propose the Quality Indicator Language (QUIL) that can be used to formally specify quality constraints on physician behaviour and patient outcomes as derived from a CPG represented in a guideline specification language such as EON, Asbru or GLIF3 [297]. QUIL has been designed to be ‘evidence-adaptive’ so that as new evidence/knowledge is acquired about
a CPG it can be quickly incorporated into the model to facilitate on-going quality assessments efforts [298]. However, QUIL requires a guideline to be encoded in an appropriate CIG representation format, which can be a significant knowledge engineering effort; also formulating quality indicators in QUIL may not necessarily be straightforward.

The Standardized Therapy Adherence Research Tool (START) [299] is also closely related to the work presented herein. START is a Microsoft Access based application that has been developed to convert records in pharmacy claims databases into a standardised format to obtain adherence related information that can then be used to design interventions aimed at improving adherence rates. START can be used to generate graphical and numerical representations of adherence for different therapeutic classes, individual agents and individual patients within a population. Adherence rates can be determined in terms of lapses in therapy as well as MPR. An important feature of START is that it can produce plots of outcome measures (BP, LDL-cholesterol and HbA1C for example) against the continuum of MPR. The application can be used to identify patients with poor adherence rates, one of the important outcomes of this research. However, I note three important aspects of ChronoMedIt over START – Firstly, drugs in START are associated with a corresponding drug class (or a group), and it is not clear whether START can account for drugs that belong to multiple drug classes when analysing adherence rates. As a result, there is no notion of a hierarchy for drug classes within START, therefore, knowledge level querying is not supported; for example, if a drug class called ACEi already exists, there is no easy way of creating a drug class called ‘antihypertensives’ and indicating that all ACEi drugs are also antihypertensive drugs. One of the possible ways to support this type of querying would be to duplicate all the ACEi drug entries and manually associate them with the new ‘antihypertensives’ group. Secondly, the level of individual patient details presented in the query results and/or reports is minimal with regard to the adherence query. The overall MPR and mean gap durations for patients are displayed, but where the medication lapse occurred, in what drug or the duration of the lapse is not displayed. Thirdly, START is designed to be a tool to determine adherence rates only, whereas ChronoMedIt is a broader quality audit framework within which adherence queries can also be formulated.

8.2.2 The clinical dimension

Medication adherence is an important aspect of treatment for chronic illness. The WHO report on Adherence to Long-Term Therapies: Evidence for Action [254] defined poor medication adherence as a critical issue for global public health, and identified five broad dimensions affecting adherence that need to be addressed by health managers and policymakers: (1) social and
economic factors; (2) health system and healthcare team factors; (3) therapy factors; (4) condition factors; and (5) patient factors. Despite the recommendations, there is little evidence that clinicians recognise patient adherence as an important factor in therapy; in fact, two recent studies that investigated patient BP control among patients who were prescribed antihypertensive medication concluded that “patients’ prior medication adherence had little impact on providers’ decisions about intensifying medications, even at very high levels of poor adherence” [272] and that “therapy intensification must be coupled with interventions to enhance medication adherence” [300].

The reasons for non-adherence are poorly understood and vary with psychosocial factors [301]. Efforts to improve adherence to antihypertensive medication can be divided into behavioural (e.g., phone reminders, packaging, dosing changes, social assistance), educational interventions (written or verbal) or combined approaches. One meta-analysis found no one approach was superior to any other but there was a trend to improved adherence with a combined approach [302]. Reminder packaging for improving adherence to long-term medication was the subject of a Cochrane review where the authors found that the intervention improved the percentage of medications taken [303]. Interventions for improving medical adherence in primary care (specifically) found patient education alone was insufficient and the evidence for complex interventions involving social support, education and reminders were inconclusive [304]. It is difficult to judge the independent effects of individual components (e.g., nurse phone call reminders) of complex interventions. An organised system of recall and regular review may be the most successful and sustainable intervention [284].

Research has shown impressive efficacy rates of BP lowering medications for reducing CVD and renal events, yet BP control levels are suboptimal at around 35% [305] and poor adherence to prescribed antihypertensive medication is often attributed this [306]. Beyond BP control, taking prescribed medications is one of the key factors to glycaemic control for many patients (especially with type 2 diabetes). A previous study reported that less than 15% of diabetic patients on single-agent therapy maintained good adherence during a one-year period, while patients on multi-drug therapy for diabetes experienced even lower levels of adherence [307]. The clinical significance of non-adherence is well established – a study reported that approximately 20% of diabetic patients were non-adherent (i.e., MPR <80%), which correlated with them having higher systolic and diastolic BP, low-density lipoprotein (LDL) cholesterol, and HbA1C, and was associated with greater risk of all-cause hospitalisation and all-cause mortality [308, 309]; after analysing refill adherence to antihypertensive medication, Heisler et
al. [272] reported that a total of 42% of elevated BP events were preceded by poor refill adherence; and Wei et al. [310] concluded that good adherence to statin treatment was associated with lower risk of recurrent myocardial infarction. As such, it has been recommended that all patients should be evaluated for medication adherence, and if non-adherence is determined to be a problem, it should be addressed so that patients can receive the full benefit of medications [308, 309]. The framework presented herein has potential to contribute to this adherence assessment function; i.e., to be used as a tool to identify specific patient cohorts to be contacted and address the non-adherence issues.

Within ChoronoMedIt, a refined definition for MPR has been used (see Section 5.3). Most previous studies only considered patients on monotherapy [256, 258-260] in an attempt to simply adherence calculations [259], which as a result, were likely to exclude patients with greater disease severity [256]. The computation ChoronoMedIt uses for its adherence calculation can handle not only patients on monotherapy, but also complex cases where patients are on multiple agent therapy; used to identify patients who have received no medication at all during a given period, as well as identify newly classified patients who show poor adherence. An important property of this MPR is that its boundary conditions are well defined. Some studies [253, 259, 311] use a variable denominator in their MPR calculation so that only the duration between the first prescription (or dispensing) and the last prescription is considered, however this requires a minimum of two prescriptions. There is an issue even with this approach however - for example, in Figure 5.2, if only prescriptions Pr3, Pr4 and Pr5 were present and such an algorithm was used, MPR will be greater than 100% (MPR = 180/125%) due to simple duration addition, whereas it is clear from the timeline that the patient had medication only for 90 days out of the 125-day period. Other studies have explicitly accounted for oversupply of medication (as that occurs during the overlap period between Pr2 and Pr3/4 in Figure 5.2) but a disadvantage of this approach is that medication lapses can be shadowed due to stockpiling. Oversupply or drug stockpiling has only a small impact [256] and it is reasonable to think that this is possibly due to the few number of high adherent patients who throughout the evaluation period collected their next prescription just a few days prior to running out of current medication. Further, accurately accounting for oversupply of medication is not always straightforward - in Figure 5.2, if Pr2 and Pr3 were for the same drug, it would imply that the 5-day oversupply needs to be added to the duration of Pr3. However, Pr4 will then become an embedded script with a shorter duration than that of Pr3. This will complicate the adherence analysis process (e.g., what is the overall antihypertensive MPR based on Pr2/3/4 with stockpiling?) and require various rules to precisely define how adherence should
be calculated. In a recent study by Vink et al. [312], overlapping prescriptions were handled by shifting the overlapping prescription(s) forward, clustering prescriptions at drug class or therapeutic level. This was based on the premise that medication may be taken sequentially, however the current position is that when scripts such as Pr3 and Pr4, or even Pr5 and Pr6 are prescribed together (where all prescriptions were for antihypertensives), these medications need to be consumed concurrently, and therefore should not be considered as covering a duration equal to the individual duration of the overlapping script. Following a similar desire to refine MPR for computation, researchers developed a new refill-based adherence algorithm called ReComp in [313]. However, ReComp was suggested as a more reliable adherence measure for short evaluation periods of around 90 days.

Past studies have investigated medication adherence rates using various techniques [253, 314] including, analysing prescribing data [312], pharmacy dispensing records [257, 315], patient self-reports [316, 317] and prescribing-and-dispensing records combined where the records are centrally stored (utilising the US Department of Veterans Affairs system) [272, 300]. However, limited studies have directly comparing adherence based on prescribing data with adherence based on pharmacy dispensing data where the prescribing and dispensing systems operate relatively independently (i.e., open-loop healthcare system). Irrespective of the way of measuring adherence, it is important for a physician to consider the concept of adherence as part of the overall chronic condition management plan, and therefore the study discussed in Section 7.7 is important as it discusses how well the adherence a clinician can ‘see’ based on their practice’s prescribing records actually aligns with the medications that were dispensed to the patient. The study showed that medication possession level as estimated from prescriptions was observed to provide 81% PPV and 76% NPV for dispensing based non-adherence and prescription based adherence was observed to be a far more powerful predictor of dispensing based adherence than other factors, such as demographics or socioeconomic status. This study based on NHI matching provided a simple and robust means of matching patients between general practice and community pharmacy. Previous research investigated success rates of linking prescribing records to dispensing records based on a combination of patient characteristics, such as gender, year of birth and postal code, and prescription characteristics, including prescription date and ATC-codes; this resulted in a lesser dispensing-to-prescribing data linkage proportion of 64.8%. Having the ability to match on NHI number, provides the more accurate indication of the value of prescribing data to indicate adherence in dispensing.
With respect to the study related to antidepressant adherence (Section 7.8), the results indicate that poor adherence rates were 39% and 68% based on prescribing and dispensing, respectively. The increase in non-adherence from dispensing as compared to prescribing is unsurprising; for example, Bailey et al. [318] report that refill failure occurred in 33% of refill opportunities. Moreover, the observed non-adherence rates in the study are consistent with the high rates of non-adherence other researchers have reported. Using a 10-day grace period for maximum lapse duration allowed, Bambauer et al. [319] reported that around 75% of the patients showed poor persistence in dispensing data for antidepressants while the results here (using a less stringent 15-day threshold) indicate that 68% of the patients showed poor adherence. Cantrell et al. [277] reported that approximately 57% of patients were non-adherent to antidepressant therapy using a six-month evaluation period, while Doesschate et al. [320] reported antidepressant non-adherence rates ranging from 39.7% to 52.7% with a mean of 46.5% over 2 years. All these studies indicate that non-adherence to antidepressants is a major issue and therefore, related barriers need to be investigated and addressed accordingly.

As discussed in Chapter 7, ChoronoMedIt has been used in several clinical contexts. The results indicate that a high proportion of the hypertensive and diabetic population (a particularly high-risk group) have deficiencies in their BP control care process that can be detected from analysis of electronic prescribing records. Also, our collaborating GPs indicate that some difficult/complex patients with hypertension and diabetes routinely get sent to a ‘difficult hypertension clinic’ (such as the one described in [321]), but the only change in therapy the specialist had to make was prescribe a diuretic or a statin [321, 322]. Specialist care is more expensive than primary care and this type of issue can be identified at the GP level (using a slight variation of the audit criteria discussed here). Other than the higher costs of referring to a specialist, there is the delay in optimisation of patient care; and it involves valuable specialist time, which ideally should be spent seeing ‘truly difficult’ cases.

Chapter 7 also discussed results from two practices where two QOF indicators were examined in the NZ context. The results showed that they would perform well on BP5 and very well on DM12 which need a 40-70% range and a 40-60% range respectively to receive the QOF points, and thereby receive incentives for managing the patients well. The observed rates are somewhat comparable to ranges reported by other researchers – a British study reported that satisfying BP5 was 83% by 2005 [323] while this study results are in the range from 65-66% for the two practices considered; the two practices considered here had a predominantly Maori/Pacific population, and there have been studies [293] indicating that such populations have inherently
higher CVD risk which may have been a contributing factor to the lower observed rates that the British study. Similarly, another British study reported that ranges for satisfying DM12 have gradually risen from 44.3% in 2002 to 70.7% in 2007 [219] while the DM12 satisfaction rates here were in the range from 59-62%. Studies have also reported that BP control has improved in the UK [222, 289] since the QOF was introduced in 2004 and the lack of a similar scheme in NZ may be a reason for the slightly lower rates of compliance with BP5 and DM12. An important result of this analysis (see Table 7.11) indicates percentages of patients failing at least one of the three indicators of interest while satisfying BP5 or DM12. Both the practices have achieved rates ≥65%, close to the upper threshold (which is 70%) of BP5 and would qualify (if they were British practices) for most of the 57 points allocated and subsequently receive financial incentives for managing hypertension patients well; however, as Table 7.11 indicates, 29-38% of these patients also satisfied at least one of the three interval-based indicators of suboptimal BP management. Similarly, both the practices have achieved a rate ≥59% for DM12 and would receive almost all the 18 QOF points. Despite being eligible for nearly the maximum financial incentives for DM12, 31-54% of the patients satisfying DM12 also failed at least one of the three interval-based indicators. These results indicate that considering other aspects of BP control is important to ensure high quality of quality indicators used to determine patients on optimal BP control. If financial incentives are to be attached to quality indicators related to BP control, they should be reliable indicators for the quality of BP management throughout the period in question.

The specific interval based criteria that have been discussed here focus on evaluating clinical outcomes with respect to time intervals (rather than points in time). Several other studies have also proposed such indicators [185, 324]; for example, a set of systematically developed primary care quality indicators for hypertension included the ‘Percentage of patients with an average systolic BP >160 mmHg and/or a diastolic BP >100 mmHg, as determined on at least three separate visits, who have a diagnosis of hypertension recorded’ [185]. It has been shown that assessing BP control based on a single measurement and/or a single visit is unlikely to be reliable, but BP considerations over time results in significantly fewer patients achieving targets set forth by guidelines [324]. Therefore, the present proposition is that although current QOF criteria are important, there is an opportunity to extend this framework to include criteria relating to BP measurements over time and perhaps to other process measures such as MPR.
8.3 Scope/Limitations

General practice PMS data reflects the quality and ‘culture’ of the GPs practicing within a given practice. Therefore, an important limitation of all the studies discussed within the scope of this thesis is that what the framework reveals is based on what is recorded in the patient EMR, which may not necessarily be complete. It is reasonable to believe that the prescribing record is complete since it is required to prescribe electronically in NZ; however, entering classification and measurement details into the EMR is to some extent at the discretion of the GP. Within the two practices we worked with, almost all the patients who had high BPs had a hypertension diagnosis indicating that the GPs were diligently entering this information into the EMR, but this may not be generalisable across practices. Further, there could be occasions where BP is measured (for patients having a ‘hypertension’ diagnosis), but not always entered, especially when patients have non-elevated BPs.

An aspect that has gradually evolved as an important use of the framework is identifying patients with poor medication adherence profiles. This identification process is based on general practice prescribing data. As such, the specific cases are those implied by the doctor-patient interactions within a particular practice. Therefore, the data provides no guarantee that prescribed medication has been dispensed; even with the studies where dispensing data was available, neither provides assurance that the patient is in fact consuming the medication as directed. Further, medication sharing among family members could be an issue as is medication wastage, both forms affecting the computed medication adherence rates. Prescribing/dispensing based adherence are imperfect adherence measures. Firstly, non-adherence based on dispensing may be overestimated where patients receive unsubsidised medicines that are therefore not captured as a pharmacy claim; although it is reasonable to believe that this is a relatively uncommon event in NZ due to the cost differences. Secondly, prescribing based adherence calculations may overestimate non-adherence in patients who receive a prescription from a prescriber other than from their ‘home’ family practice (a specialist for example). However, a main focus of the present research was based on taking the prescribing record alone as the basis for identification of patients in need of quality improvement action; and this research has shown that there is inherent merit in such analysis, at least from the GP’s perspective, as prescribing is a direct action that has been taken on a patient which is a reflection of GP adherence to evidence based guidelines. It is interesting that prescribing alone already uncovers such substantial cohorts for intervention; however, extension to include dispensing, or even home telemonitoring to allow more frequent biological measurement, would obviously result in superior ‘intelligence’ and is a desirable future
direction. Pending wider uptake of such solutions, however, the results indicate a large reservoir of patients who are providing clear indications of adherence problems based on poor medication possession as well as other quality indicators.

The use of only one or two general practices within the scope of data analysis can be seen as another important limitation. The results may be biased with respect to the relationship of those staff to their patients, or by the nature of the patient population vis-à-vis another to which one may wish to generalise. In particular, results may be expected to vary in jurisdictions that provide lesser medication subsidies (the NZ patients pay only NZ$3 per three months supply of each drug for the subsidised medications analysed). Hirth et al. [325] found that cost-related non-adherence ran highest in the US (out of 12 nations examined) at 17%. Thus it is realistic to expect that dispensing based adherence will be lower vis-à-vis prescribing based adherence in such healthcare systems, and particularly among patients at the lower socio-economic levels.

Preliminary result validation discussions with collaborating GPs indicate some suboptimal management of patients occur due to GP-system interaction errors, as opposed to patient related factors. For example, all the data extractions were performed using the MedTech32 system, the most widely used PMS system in NZ, and within this PMS GPs have to mark long-term medications as ‘Long-Term’; if a GP has not marked a certain medication as being long-term, and then at a later date decides to prescribe using the “prescribe all long-term medications” function of the PMS, the medications not marked as ‘Long-Term’ will not get prescribed. Some of the reports presented to practices had highlighted certain patients as having diabetes and not being on ACEi/ARB for over 30 days, but before a patient was recalled, the practice nurse had reviewed the patient EMR to note that this medication lapse occurred due to a GP-system interaction error rather than a patient related issue. This has been a positive (unexpected) side-effect of this type of reporting; however, the reports do not account for these types of errors at the moment.

In terms of the implementation of the computational framework, it has been driven by clinician requirements, and as such implements only the most general forms of the eight audit criteria that were established. The implementation has been generalised as far as possible to handle any type of drugs or drug classes, including combination drugs (opposed to only antihypertensives as required by clinicians); any domain-level diagnosis or specific diagnosis codes; specify any measurement, such as BP, eGFR, uric acid and so on; any number of consecutive measurements (vs. three BPs); and query using any comparison operator. These classes of queries reflect the
scope of the initial audit criteria and therefore it is not possible to formulate criteria outside the scope of these. As a result, the current implementation is tailored to medications that are taken regularly, in discrete, uniform doses and the framework is not at present sufficiently fine-grained to assess treatment where the patient (or the clinician for that matter) adjusts the dose in response to conditions, as with adjustment of insulin dosing based on blood sugar. Further, it does not account for controlling outcomes such as BP by means such as managing a patient’s diet and/or exercise.

8.4 Future directions

My supervisor and I have received positive feedback from our collaborators on ChronoMedIt’s current capabilities and there are several intended future directions. Our discussions with practice clinicians indicate that CVD risk (commonly termed CVR) is an important concept when identifying ‘at-risk’ patients, for example, patients with diabetes who have CVR >15% with BP >130/80 mmHg. ChronoMedIt’s Criteria Model currently does not support this type of temporal relationships which needs the BP measurement to be after a high CVR event and the required diagnosis. However, this should be a fairly simple extension to the CommonMeasurementCriteria class (see Section 5.4) where the new derived criterion class will need to specify the minimum required CVR threshold. For some patients the clinicians calculate CVR (either manually, or via PREDICT - a NZ based decision support tool that is often integrated into MedTech32 to determine CVR and provide recommendations to the GP and patient [326]), but currently the data extract does not contain CVR information. Extending ChronoMedIt to either internally calculate CVR based on a suitable algorithm or extending the data extraction protocol (with ethics approval) to include CVR information may be a potential solution for this.

NZ is not unique in having high-quality prescribing data at the practice level that could support the type of quality improvement efforts discusses herein. This is true on a national scale for a number of countries, including the Netherlands, the UK and Australia [327]; and in some cases, as in Australia, the prescribing data will be far more readily accessible to practices for quality improvement efforts than the dispensing data which is collected as part of a national reimbursement scheme. Moreover, there are other large databases that capture prescribing data (e.g., the General Practice Research Database [GPRD] [328] of anonymised longitudinal medical records from primary care that collects data from around 488 primary care practices throughout the UK) which would be suitable for epidemiologic analyses of adherence based on prescribing.
As such, possible collaboration with relevant institutes from such counties is currently being considered.

As discussed previously, the current proposition is that prescribing data alone can uncover many deficiencies in terms of patients being on suboptimal therapy. Additional sensitivity in detecting medication non-adherence could be provided by an e-pharmacy network, such as that deployed nationally in Denmark [329], where primary care physicians (as well as patients) can identify dispensing against a specific prescription via a central database. The prescribing-dispensing matching studies discussed in Chapter 7 indicate that a prescription is generally followed-up by a timely dispensing event; given that this first-fill rate far exceeds overall adherence, it appears that it is in the refills that patient non-adherence is most commonly first revealed. With an e-pharmacy network it would be possible to detect this earlier breakdown in adherence and provide timely follow-up (e.g., a phone call). Moreover, some patients show high adherence via dispensing and yet show non-adherence in the prescribing records of the practice (reducing the PPV of prescribing based MPR on dispensing based MPR for non-adherence). This is understandable in that, without a national network, the patients could have received prescriptions for a period of time from other providers, including specialists and hospital outpatient settings. The reimbursement network present in NZ is not currently suited for extension to perform this pharmacy information function since reimbursements are collated and processed in batches with potentially lengthy delays of up to 60 days. More complete monitoring of the prescribing, dispensing, consumption cycle would be desirable; for instance in NZ, linkage of prescribing to dispensing will be greatly facilitated if the NZ Health Informatics Standards Organisation (HISO) draft e-pharmacy specification [330] is adopted.

Section 7.9 discussed the details of an ongoing feasibility trial where ChronoMedIt is being used to identify patients with poor adherence to antihypertensive medication. Based on this feasibility study, a current summer medical student intern who has been analysing EMR data has identified cases where the patient has been adherent and BP not controlled, however, there has been no therapy intensification for lengthy periods of time, around a year in some cases. This finding is consistent with other literature where it has been reported that medication adherence as well as suitable therapy intensification need to be considered concurrently for successful BP control [300]. Considering therapy intensification alone may not be sufficient, as it is possible that patients whose therapy was intensified despite non-adherence could experience episodes of hypotension (i.e., BP is too low), a commonly raised concern in such a situation [288]. As such, there are plans to extend ChronoMedIt to include dose related information as
well to determine episodes of therapy intensification. Providing therapy intensification details combined with (non-)adherence information will assist clinicians to have a more complete view of a patient’s hypertension management, and develop care plans accordingly.

Beyond expansion of the data analysis framework per se, an immediate future direction of this work is to develop a web service where ChronoMedIt functionality can be provided for other applications to use. There are also considerations around developing a Clinical Document Architecture (CDA) [331] based specification for ChronoMedIt’s patient data structure so that PMSs or other data repositories with CDA support could use the auditing functionalities of ChronoMedIt.

Another immediate future work program is to investigate the issues/reasons associated with those patients who are on suboptimal therapy. The framework does not inherently provide any solution with respect to how the information is to be used by a practice to address the issues. Currently, a research team led by my supervisor is preparing an application for the NZ HRC to extend the ongoing feasibility study to a larger RCT which will assess the effectiveness and uptake of PMS-based monitoring of prescribing to support systematic practice-nurse telephone outreach for improving adherence (measured using MPR), BP and renal function over 12 months among patients on antihypertensive medication. The expected timeline of the trial is shown in Figure 8.1.

![Timeline of the randomised controlled trial](image)

*Target to complete all initial visits in first 6 months, scheduled based on staff availability and patient’s current prescription expiry

**Follow-up contacts continue for 1 year from initial contact for each patient. Intervention ramps down in last 6 months, mirroring initial 6-month ramp-up.

Figure 8.1: Timeline of the randomised controlled trial
The study is designed as a cluster randomised controlled trial where high needs practices (i.e., greater than 50% of enrolled patients are Maori, Pacific or low socio-economic status) in Auckland and Waikato will be enrolled. Patients enrolled with these participating practices who have MPR <80% for antihypertensives in the previous year will form the study cohort. PMS data will be used to identify patients (via ChronoMedIt) with poor history of timely re-prescription (MPR <80%) for 12 months of practice nurse-led telephone outreach emphasising timely recall, identification of adherence issues, and review of eligibility to and participation in existing chronic care programmes. Control will be ‘usual care’. The primary outcome of this study will include MPR and systolic BP while the secondary outcomes will be assessed in sub-populations, including, eGFR among those at increased risk and HbA1c among those with diabetes. This study intends to enrol approximately 6000 patients on regular antihypertensives with an MPR <80% from 30 practices to detect a mean difference in change between intervention and control participants of 5mmHg in systolic BP over 12 months. While the study targets antihypertensive medication, it is anticipated that adherence to other cardiovascular or diabetes medication in the targeted patients will also improve. Further details are outside the scope of this thesis and will be published in the future, subject to receiving funding.

While the current focus has been on the process of clinical audit, largely within a single practice and with locally-defined and agreed reporting criteria, the concepts around aggregate analysis (for a group of practices, a region, or nationally) are not unrelated. CBG’s HealthStat [332] demonstrates the potential for similar reporting based on national-wide sampling of PMS data on short reporting cycles (e.g., weekly) and this could provide an attractive pool of clinical data for ChronoMedIt’s auditing purposes.

The developed computational framework – ChronoMedIt – shows potential to be used as a tool to improve clinical outcomes. It supports analysis of practice EMRs with consideration of various non-trivial temporal relations between prescriptions, measurements and problem classification within the context of a specified evaluation period. Further work is needed both to expand the EMR-based reporting to various other important classes of quality audit criteria for chronic condition management and to determine how to effectively utilise such reporting to improve patient outcomes.

Despite the drawbacks discussed in Section 8.3, there is room for using the proposed framework to assist general practice physicians to identify chronic patients whose health outcomes can be improved. Within the scope of this PhD research, I have demonstrated
ChronoMedIt’s ability to provide quality audit reporting primarily with regards to hypertension, but since the nature of chronic illness is much the same across the chronic disease space, with minimal changes the framework can be applied to other chronic illnesses as well. Our experience with clinicians [17] has already shown that GPs have much interest in being able to identify chronic patients who have medication lapses (and therefore low medication adherence). To this end, we are looking forward to analysing the data from the ongoing feasibility trial, as well as carrying out the RCT mentioned previously.

A recent German study [333] reported that “...practicable integrated quality management...is a significant innovation of chronic care management and an efficient way to improve [chronic disease] continuously”; in that study GPs received quarterly quality reports based on collected data with various outcome indicators, further indicating that the envisaged future directions for ChronoMedIt provide a good template for better chronic disease management.

It is herein hoped that the contribution of this thesis to the larger medical informatics community will ultimately result in improved chronic disease management where quality audit indicators are used to their fullest potential.
Appendix I – A Quality Audit Report

Prescribing Analysis for the Pasifika Healthcare
(focussing on antihypertensive prescribing)

Evaluation Period: From 09 May 2007 to 09 August 2007

I. General

1. Total number of patients (on 08 May 2007) 11459
   - Funded 6575 57.4%
   - not funded (Deceased, Enrolled Elsewhere and Rejected) 4884 42.6%
2. Age-wise categorisation of funded patients
   - < 40 years 4556 69.3%
   - 40 - 70 years 1780 27.1%
   - 70 years and above 239 3.6%
3. Number of funded patients with prescriptions 3664 55.7%
4. Number of prescriptions issued to funded patients 29433 100%
5. Number of funded patients with prescriptions with one or more classification(s) 3271 49.7%
6. Number of BP measurements 13470
   - valid 13166 97.7%
   - invalid (ie. not within range of 30-300 mmHg) or incomplete 304 2.3%
   - measured by a GP or nurse 8663
   - measured by GPs 1426 16.5%
   - measured by nurses 7237 83.5%
7. Number of funded patients classified during the period 08 May 2002 – 08 May 2007 with
   - hypertension 659 10%
   - diabetes mellitus 586 8.9%
   - hypertension and diabetes mellitus 289 4.4%
   - renal disease 157 2.4%
   - cardiovascular disease (heart failure, myocardial infarction and/or angina) 65 1%

3 This number is less than the total BP measurements since some of the initials (corresponding to the measured person were not there in the list of GP/nurse initials provided.
4 A patient can have more than one classification.
II. Antihypertensive Prescribing

II.1 Quality indicators supporting the practice

II.1.1. BP Measurements*

1. Number of funded patients with at least three BP measurement after being classified 515 78.1%
2. Number of funded patients not classified with diabetes with last BP measurement controlled (≤ 140/90mmHg) 454 68.9%
3. Number of funded patients with at least one BP measurement controlled after being classified
   - ≤ 140/90mmHg 525 79.7%
   - ≤ 130/90mmHg and classified with diabetes 243 36.9%

II.1.2. Prescriptions

4. Number of prescriptions with antihypertensive agents issued to funded patients (percentages based on this value) 5887
   - ACEi/ARB prescriptions 2327 39.5%
   - Beta-blocker prescriptions 1060 18%
   - Diuretic prescriptions 1556 26.4%
   - Non-DCCB prescriptions 159 2.7%
   - DCCB prescriptions 663 11.3%
   - “Last resort” prescriptions 122 2.1%

5. Number of funded patients prescribed with antihypertensive agent(s) prescribed at least once after classification (percentages based on this value) 517
   - ACEi/ARB prescriptions 440 85.1%
   - Beta-blocker prescriptions 195 37.7%
   - Diuretic prescriptions 265 51.3%
   - Non-DCCB prescriptions 26 5%
   - DCCB prescriptions 114 22.1%
   - “Last resort” prescriptions 15 2.9%
### II.1.3. Effective Combination Therapy*

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Number of funded patients who have been on concurrent therapy with ACEi/ARBs and diuretics</td>
<td>221</td>
<td>33.5%</td>
</tr>
<tr>
<td>7</td>
<td>Number of funded patients who have been on concurrent therapy with beta-blockers and diuretics</td>
<td>103</td>
<td>15.6%</td>
</tr>
<tr>
<td>8</td>
<td>Number of funded patients who have been on concurrent therapy with beta-blockers and DCCBs</td>
<td>46</td>
<td>7%</td>
</tr>
<tr>
<td>9</td>
<td>Number of funded patients who have been on concurrent therapy with ACEi/ARBs, diuretics and non-DCCBs</td>
<td>13</td>
<td>2%</td>
</tr>
<tr>
<td>10</td>
<td>Number of funded patients who have been on concurrent therapy with more than three antihypertensive agents</td>
<td>1</td>
<td>0.2%</td>
</tr>
</tbody>
</table>

### II.1.4. Continuity of Therapy*

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Number of funded patients with no lapse in antihypertensive therapy for more than 90 days</td>
<td>558</td>
<td>84.7%</td>
</tr>
</tbody>
</table>

### II.1.5. Drug-Problem Indication*

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Number of funded patients classified with diabetes mellitus prescribed with ACEi/ARBs</td>
<td>233</td>
<td>35.4%</td>
</tr>
<tr>
<td>13</td>
<td>Number of funded patients classified with myocardial infarction and prescribed with beta-blockers</td>
<td>128</td>
<td>19.4%</td>
</tr>
</tbody>
</table>
II.2 Quality indicators critiquing the practice

1. Number of funded patients not classified with hypertension and prescribed with antihypertensive agent(s) 253

2. Number of funded patients not classified with hypertension and with three consecutively high BP measurements (≥ 140/90mmHg or ≥ 130/90mmHg and ever classified with diabetes) 111

3. Number of funded patients classified with hypertension where hypertension is not highlighted 644 97.7%

4. Number of funded patients classified with hypertension and not prescribed with antihypertensive agent(s) within 90 days of classification 3 0.5%

5. Number of funded patients classified with hypertension and with last BP measurement not controlled (≥ 140/90mmHg or ≥ 130/90mmHg and ever classified with diabetes) 181 27.5%

II.2.1. BP Measurements*

6. Number of funded patients classified with diabetes with last BP measurement not controlled (> 130/90mmHg) 39 5.9%

7. Number of funded patients with three or more consistently high BP measurements (≥ 160/100mmHg) for over 120 days 25 3.8%

8. Number of funded patients with at least one (systolic or diastolic) high BP measurement (≥ 160/100mmHg) and no further BP measurements for over 120 days 169 25.6%

II.2.2. Lack of Continuity of Therapy*

9. Number of patients with lapse in antihypertensive medication for:

<table>
<thead>
<tr>
<th>Duration</th>
<th>Total</th>
<th>With no BP measurement in lapse period</th>
<th>With at least one high BP measurement during lapse period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>With some encounter with the practice</td>
<td>With no encounter with the practice</td>
</tr>
<tr>
<td>&gt; 30 days</td>
<td>257</td>
<td>35</td>
<td>201</td>
</tr>
<tr>
<td>&gt; 60 days</td>
<td>160</td>
<td>21</td>
<td>113</td>
</tr>
<tr>
<td>&gt; 90 days</td>
<td>101</td>
<td>13</td>
<td>65</td>
</tr>
<tr>
<td>&gt; 120 days</td>
<td>63</td>
<td>4</td>
<td>42</td>
</tr>
</tbody>
</table>

10. Instances where time between prescriptions exceeded the desired duration by more than:

- 3 days 2995
- 5 days 2875
- 10 days 2484
- 30 days 1816
II.2.3. Drug-Drug Interactions*

11. Number of funded patients prescribed with anti-asthmatic drugs and beta-blockers  
   14  2.1%

12. Number of funded patients who were on concurrent therapy with beta-blocker and Verapamil:
   - classified with atrial fibrillation and/or flutter  
     0  0%
   - not classified with atrial fibrillation and/or flutter  
     0  0%

II.2.4. Drug-Problem Interactions*

13. Number of funded patients classified with asthma and prescribed with beta-blocker(s)  
   4  0.6%

14. Number of funded patients classified with gout and prescribed with thiazide(s)  
   42  6.4%

15. Number of funded patients classified with diabetes mellitus and on monotherapy with DCCB  
   2  0.3%

16. Number of funded patients classified with gout, not on Allopurinol and prescribed with thiazide(s)  
   18  2.7%

II.2.5. Monitoring*

17. Number of funded patients prescribed with thiazide(s) and with serum uric acid levels > 0.42mmol/l  
   127  19.3%

Report Generation
Start time: 22 November 2007 03:29:08 PM
End time: 22 November 2007 03:29:46 PM
Time taken to generate report: 38 seconds.

* Unless stated otherwise, all patient percentages are as a percentage of funded patients classified with hypertension (and not as a percentage of total funded patients) and all prescription percentages are as a percentage of total antihypertensive prescriptions (and not as a percentage of total prescriptions). All patient and prescription figures are for patients classified with hypertension.

†A lapse in antihypertensive therapy is defined to commence when all antihypertensive medications if taken as directed from the day of prescribing, should have run out. Brief lapses are expected and may not be problematic where the patient has retained some prior supply.
Thesis Related Publications

Refereed journal articles

Lead author


Contributing author


Refereed conference papers

Lead author


Contributing author

References


144. SNOMED Clinical Terms. Available from: [http://www.snomed.org](http://www.snomed.org) [cited 22 June 2009].


191. *Pan-Canadian Primary Health Care Indicators, Report 1, Volume 1*. 2006. Ontario: Canadian Institute for Health Information

192. *Canadian Institute for Health Information Primary Health Care Indicators Electronic Medical Record Content Standards, Version 1.1*. 2009. Ontario: Canadian Institute for Health Information


249. Booth, G.L. and Hux, J.E., "Relationship between avoidable hospitalizations for diabetes mellitus and income level," Archives of Internal Medicine, 163(1): pp. 101-6, 2003


