An Adaptable Long-term Condition Workload Prediction Model for Primary Health Care

Manjula Devananda

a thesis submitted for the degree of

Doctor of Philosophy

at the University of Otago, Dunedin,

New Zealand.

30 June 2018

Abstract

The primary health care (PHC) system must manage the growing demand for care due to patients with long-term conditions (LTCs) such as diabetes, hypertension and asthma. Population-based care can help address this workload management problem, specifically by enabling a shift from a reactive to a proactive patient management approach. However, current PHC systems lack the ability to provide population-based care. This thesis presents a tool to predict the future workload generated by a population of patients.

We use a rule-based system, for its modularity, flexibility and the automated modifiability behaviour, to develop the care pathways as rules that, when given the patient data, would simulate the patient visits for the upcoming year (since some start date). It is assumed that the GPs follow best practice and their patients adhere to their plan-of-care, making visits to the medical practice on their scheduled LTC appointments. Then, these visits are aggregated to a population-level as a count of appointments per week from these LTC patients, referred to as the workload to be managed within the capacity of the practice. Knowing this predicted workload, the PHC organisation can then plan and deliver care accordingly. In this thesis, we also explore using seven what-if scenarios the impacts of alternatives in practices and evaluate the strategies to address them. We then propose the use of Bayesian inference in our workload prediction model, in order to incorporate the variation in patient visits due to the impact of renewal of their LTC prescriptions.

This work is done in collaboration with BPAC, a non-profit organisation that promotes best practice for primary care within New Zealand. The collaborations on this work were in the form of

1. health data from a medical practice;
2. the knowledge base to understand the primary health care domain, and the practical issues at a medical practice.

Approval for this research using the anonymised patient data provided by BPAC has been given by the University of Otago Human Ethics Committee (Health).

We follow the design science research (DSR) methodology to develop our adaptable best practice based workload prediction model for a PHC due to its LTC patients. From a DSR perspective, we developed a construct called the three-layer LTC PHC construct, and a process called the encounter-based unfolding plan-of-care process, which are used to build our artefact: the adaptable best practice based workload prediction model. In DSR, much emphasis is on communicating the developed artefact or model to wider community and the feedback guides further improvement of the model or the artefact developed. We followed seven iterative cycles to incrementally build the rule base, and the feedback served as a guide to improve the simulation capability of our ABP-WPM. Apart from the feedback from the collaborator on this work, feedback was also collected through informal meetings with the care providers of a medical practice at Mosgiel, the executive members of two different PHOs (one in the North Island and the other in the South Island of New Zealand) at various stages of development of this artefact. The artefact developed was also communicated to the research community through two publications.

In the process of developing this population-level workload prediction model, we identified shortcomings (for example, the LTC status of a patient is not explicit) in the current health data models and in the PHC data shared with us, which is needed to support a population-level workload analysis. We therefore developed a patient information data model that makes this information explicit.

Acknowledgements

This journey would not have been possible without the support of my family, professors and mentors, and friends. The gratitude, respect, admiration and love I have for my mother-in-law Radhamoni, my husband Devanand, and our maid Vimala, who took the inimitable responsibility and gave all the love and happiness to my wee boy Dyutit (4 year old now), is beyond words. I must say I am indebted to them for bringing up my son just as I would have if I were around him. Although it is very hard for a mother to be away from her child, their love and care made the pain bearable.

I am extremely grateful to my supervisors Prof. Stephen Cranefield, Prof. Michael Winikoff and Dr. Hywel Lloyd for their unrivalled support and guidance throughout this journey. I enjoyed and learned a lot from them, especially I loved them challenging my thinking by helping me question assumptions and view issues from multiple perspectives. Above all, I am very grateful to them for giving me a feeling that I am part of a great team.

I am more than grateful to the Best Practice Advocacy Centre (BPAC) for providing the anonymised dataset required for this study. I sincerely appreciate the comments and suggestions received from the members of the Pinnacle Midlands Health Network (a PHO of North Island), the Well-South Primary Health Network (the PHOs of Otago and South Island) and the health professionals of Mosgiel primary health centre. I am grateful to Matthew Schofield and Matthew Parry from the Department of Mathematics & Statistics of our university for their advice on the Bayesian model developed in this work.

I must thank all the faculty members and colleagues of my department (Information Science) for providing me with feedback on this both formally

and informally during our research group meetings. Special thanks to Associate Prof. Tony Saviramuthu for our (constructive) informal chats and long coffee breaks. Also, special thanks to our department manager Stephen Hall-Jones and the department administrators Heather and Gail for their support and services (including printing of this thesis) received during this study period.

I am grateful to the Graduate Research School and the Scholarships office for the financial support granted (Doctoral scholarship) for the past three years. I also take this opportunity to thank the Collaborative On-Demand Learning (CODeL) team, HEDC department, librarians of our Science and Health Science divisions and the postgraduate student community of Commerce division for providing help and support through workshops and other gatherings.

I am grateful to my friends, especially Stephy Simon for all our long drives, bush walks and of course the food. To my teachers and friends of Natyaloka School of Indian Dance, for adding dance and music to my life and making this journey more enjoyable. To my Lovelock house mates, thank you for listening, offering me advice, and supporting me through this entire process. To my friends and relatives scattered around the globe, thank you for your thoughts, well-wishes/prayers, phone calls, e-mails, texts, and being there whenever I needed a friend.

I take this opportunity to thank the Almighty God, for blessing me with the right people (which includes my supervisors and friends) at the right time who gave me the right advice to be a happy person (I smile always).

Contents

List of Tables

List of Figures

List of Abbreviations

ABP-WPM	Adaptable Best Practice based Workload Prediction Model
ABS	Agent-Based Systems
BPAC	Best Practice Advocacy Centre
CCA	Care Continuum Alliance
CCM	Chronic Care Model
CDSSs	Clinical Decision Support Systems
CIGs	Computer Interpretable Guidelines
DES	Discrete Event Simulation
DSRM	Design Science Research Methodology
EBC	Evidence Based Care
GP	General Practitioner
II	Information Interpretation
KP	Kaiser-Permanente
LTCs	Long-Term Conditions
MCNs	Managed Clinical Networks
PBC	Population Based Care
PCA	Primary Care Advisor
PCC	Patient Centred Care
PH	Public Health
PHC	Primary Health Care
PHO	Primary Health Organisation
QDM	Quality Data Model
RBD	Rule-Based Development
SC	Simulation Capability
SD	System Dynamics
TBC	Team Based Care
WHO	World Health Organisation

Glossary

Active medication	A medicine prescribed for a patient within four months of the date of appointment of the patient.
Best practice guidelines	Clinical guidelines reviewed to meet the health care management needs locally within a country, region or a medical practice.
Care pathway	Focuses on a specific LTC management needs.
Care plan	Focuses on a specific LTC management need in an individual.
Longterm Condition	The health condition of a patient that is longterm and can be managed through timely interventions; e.g., diabetes, hypertension and asthma.
Multi-morbidity	Coexistence of more than one LTC in a patient.
Plan-of-care	Addresses multi-morbidity management needs in a patient.
Population-based care	Planning and delivering care for a group of patients who share common health care needs.

Chapter 1

Introduction

Worldwide, long-term conditions (LTCs) such as diabetes, hypertension and asthma are becoming more common as the population ages (He et al., 2016; Ministry of Health NZ, 2016b). The New Zealand Burden of Diseases, Injuries and Risk Factors Study found that older people will live longer with multi-morbidity[1] and associated health care needs (Ministry of Health NZ, 2016c). Consequently, costs for care and the challenge of providing care to patients with LTCs are expected to grow significantly in the next 20-30 years (Mabotuwana and Warren, 2010; Mays, 2013; National Health Board NZ, 2010). This thesis addresses managing this growing demand for care due to LTC patients at a primary health care level.

A primary health care (PHC) system is usually the first point of contact with the health system of a country (WHO, 1978). The Ministry of Health NZ (2017i) describes a PHC as "*the professional health care provided in the community, usually from a general practitioner (GP), practice nurse, pharmacist or other health professional working within a general practice*". Likewise, most countries usually aim to provide the general health care services through a general practitioner (GP) or a practice nurse at a PHC centre (Montague, 2014).

LTCs are defined as "*any ongoing, long-term or recurring conditions that can have a significant impact on people's lives*" (Ministry of Health NZ, 2017c). Due to the LTCs becoming more common, recently, in addition to providing general care, the focus of PHC systems is turning towards prevention, early detection and well-management of LTCs in a patient (Ministry of Health NZ, 2016b,d; WHO, 2016b). As most LTC patients live with multi-morbidity, more complex care is needed to meet their LTC management needs (Johnson, 1997; WHO, 2016c). So, ideally, multiple care providers

[1]Multi-morbidity is the coexistence of more than one LTC in a patient.

should provide care, following a care planning process that considers all the LTCs in a patient, and results in a shared *plan-of-care*[2] for the patient (Burt et al., 2014; Wagner, 1998; WHO, 2016c). A plan-of-care is thus expected to address the health care needs of all the LTCs, including shared goals and benefits of interventions by various care providers, as required for the patient (Burt et al., 2014).

In order to meet the demand of care for (multi-morbid) LTC patients, the PHCs need to plan. Planning is required for individual patient care, but LTC patients benefit better from planning for a *population* of patients (Wagner, 1998), which has led to more attention being given to *population-based care*. Population-based[3] care considers care for a group of patients who share a common trait (Wagner, 1995). This common trait could be sharing the same GP, being diagnosed with the same LTC, or following a similar plan-of-care (WHO, 2005, pp. 45). Population-based care, according to Wagner (1995) "*uses guidelines, and epidemiologic data and techniques to plan, organize, deliver and monitor care to specific clinical sub-populations such as diabetics*". Population-based care can thus help a PHC system to address the LTC management needs of a group of LTC patients, and also, to equip the PHC system dynamically, i.e., according to the growing demand of care due to these LTC patients (Mays, 2013; Ministry of Health NZ, 2017k; WHO, 2016c). This planning for a population of LTC patients can also allow a shift from reactive, episodic patient management to a proactive, continuous and systematised care.

1.1 The PHC workload management challenge

Currently, the PHC system is a patient-initiated, reactive system (Montague, 2014). Moreover, the PHC system is designed for acute, shorter period appointments (Wagner, 1998). Usually, GP appointments are fixed 8, 10 or 15-minute slots (Wallace et al., 2015; WHO, 2002). However, LTC patients require longer appointments to address the health care needs of multiple LTCs (if any) in them (Johnson, 1997). Therefore, the growth in the number of people with LTCs will impact adversely on the performance (both clinical and financial) of PHCs. Hence, there is a need to shift this current practice from a patient-initiated, reactive care setting to a proactive, systematised PHC setting for better, efficient and effective chronic care (Ham, 2010; Ministry of

[2]Chapter 2 provides a detailed discussion of plans-of-care, as well as related concepts such as care plans, and care pathways.

[3]In this work, population-focused, population-based and population-level are used interchangeably to relate to care targeting a group of patients.

Health NZ, 2000). A population-based approach will help to make this transformation systematic and smooth (Pines et al., 2015). However, a PHC system faces three major, interlinked, challenges that hinder planning care for a population of patients.

1. LTC patients account for the vast majority of appointments within the PHC system (McPhail, 2016). Furthermore, their multi-morbidity requires longer appointments (Johnson, 1997), which in turn, adversely affects the overall efficiency of a medical practice.

2. There is a gap between the demand for care and the capacity of a medical practice to handle the challenges due to LTCs (Hefford, 2006; Ministry of Health NZ, 2000; Townsville Mackay Medicare Local, 2012). Moreover, government health strategies such as Victoria's International Health Strategy 2016-2020 (Department of Health & Human Services, 2016), New Zealand Health Strategy 2016 (Ministry of Health NZ, 2017h), and the Public Health Strategy by the Department of Health Nunavut (2018), that aim to improve LTC patient outcomes through integrated care, unfortunately have adverse affects on the workload of health professionals (Stokes et al., 2017).

3. Usually, due to various barriers in a PHC context, both primary care providers and patients tend not to adhere to clinical guidelines (Fischer et al., 2016; Haynes and Haines, 1998; Overington et al., 2014). Moreover, though studies (Chaudhry et al., 2006; Woolf et al., 1999) show that adherence to clinical guidelines improves the quality of care and enables patients to manage their LTCs better, intervention by multiple care providers often leads to fragmented LTC care (Stokes et al., 2017; WHO, 2016c).

1.2 Research questions

The increase in the number of patients with LTCs poses a major challenge to managing the workload at a PHC centre. Population-based care can help address this workload management problem, specifically by shifting from a reactive to a proactive patient management approach. However, managing patients proactively requires planning tools that can plan, organise and deliver care (Wagner, 1995). Currently, PHC systems lack such planning tools (Ministry of Health NZ, 2017k). Hence, for this research, the problem is identified as to "*predict the workload,* ***if a medical practice followed***

the best practice[4] *for its LTC patients. The PHCs also need a tool to explore various ways to manage this workload within the capacity of a practice, which may also include variations in the health policies at a practice.*"

Specifically, this thesis addresses the following research questions:

RQ1 **Given a medical practice that follows best practice plans-of-care**, what model(s) can be used to predict the population-based care workload? This question specifically focuses on the workload that is expected to arise from the demand of care to meet the LTC management needs in a PHC context.

RQ2 How can this predicted workload be analysed in various what-if scenarios? For example, if the less severe patients are attended by nurses, or if a new model of care is adopted for their LTC patients, what is the change in workload of the GPs?

RQ3 How can the impact of various health policies be studied at an organisational level? For instance, what is the impact on workload if all LTC patients at the practice should have an annual CKD screening?

(RQ1) Workload prediction

A PHC centre provides "general" care, which includes care for acute cases, immunisations, LTC management needs and acute exacerbation of LTCs in a patient (WHO, 1978). Attending to these health care needs, the required interventions could be events such as clinical reviews, observing lab measurements and prescribing medications, that contribute to the workload on various care providers at a PHC centre. We address this workload from three dimensions.

1. Stratify LTC workload from other cases; We focus on the workload that arises specifically from LTC management in a patient population at the PHC centre.

2. Focusing on the workload of the GPs; Assuming that currently the LTC patients are attended by GPs, we focus on the workload on the GPs.

3. Aggregating the number of clinical reviews (hereafter referred to as recalls) of LTC patients for their LTC management needs, to present weekly workload figures;

[4]Best practice guidelines are clinical guidelines reviewed to meet the health care management needs locally within a country, region or a medical practice (Johnson, 1997).

We assume that GPs follow the best practice to manage their LTC patients, hence the recall decision for a patient is driven by care pathways (see Section 2.1.1) applicable to them.

(RQ2) What-if scenarios

Knowing the upcoming workload can help to plan resources accordingly (Heroman et al., 2012). What-if scenarios explore various options to manage the predicted workload at a PHC practice. A few examples include shifting low risk patients from a GP to a practice nurse which would obviously, make the GP more available for other patients. This analysis would justify the measures taken to manage the LTC workload efficiently within the capacity of the practice. This would also help the organisation equip itself with the right mix of various care provider roles to meet their LTC patients' needs. A change in policy or practice could have both clinical risks (i.e., related to the quality of care delivered) and financial risks (i.e., the financial impacts) on the practice. In our what-if scenarios, therefore, we also explore the impact of financial risks for the practice.

(RQ3) The impact of health policies

Most government health strategies are implemented at a national level. For example, recently, the Ministry of Health NZ (2016a) implemented their "Living well with Diabetes" plan to address the prevalence of diabetes among the New Zealanders. However, currently medical practices lack the ability to predict and understand the impact of a health strategy at an organisational level - the volume of increased access to health care services and its impact on health professionals is unknown. This thesis presents a model that enables the impact of such policies for a PHC centre to be simulated. This would help the organisation to review the best practice guidelines and re-adjust its policies accordingly.

While RQ 2 and RQ 3 are closely related, they address two different aspects. Specifically RQ 2 addresses aspects within the medical practice where there is more autonomy for the practice to tackle the workload changes. While RQ 3 addresses specifically the impacts of national level strategies where the medical practice has less autonomy but can address the impact in some way within its capacity.

Having the care pathways depicted in an "If-Then" format (Alther and Reddy, 2015), we use a rule-based approach in this work to simulate the care pathways and

patient visits to predict the workload of a GP, and to explore various what-if scenarios to manage this predicted workload. The use of rule-based models in health care domain is widely accepted.

The benefits of using a rule-based system include the following. (Minutolo et al., 2017; Shiffman, 1997; Shortliffe, 1974):

1. We can extend the rule-base by adding new rules, i.e., the modularity feature of rule-based systems where rules can be added (extended) without need for change of other rules; this feature helps us to incrementally build and evaluate our workload prediction model through an iterative cyclic approach presented in Chapter 3.

2. We have the flexibility to choose different rules based on the facts expressing the current context, i.e., the openness feature of the rule-based systems enables us to have rules for an environment and only those rules that match the specific context would fire. This feature is extensively helpful in medical systems. In our case, this feature helps us to apply a plan-of-care to a patient.

3. It enables dynamic handling of changes through rules being re-activated based on changes to facts during the execution of rules, i.e., the automated modifiability behaviour of rule based systems. This feature is essential to explore our what-if scenarios. In our what-if scenarios, we require to change the rules based on the context. For instance, in a scenario it would first allocate workload to the GPs, which later need to be re-scheduled to the nurses.

In a nutshell, this study aims to simulate patient visits for the upcoming year (since some start date). It is assumed that the GPs follow best practice and their patients adhere to their plan-of-care, making visits to the medical practice on their scheduled LTC management recalls[5]. Then, these visits are aggregated to a population-level as a count of appointments per week from these LTC patients, referred to as the workload to be managed within the capacity of the practice.

Apart from answering these research questions, this work makes two more contributions.

1. We built a workload prediction model which assumes that the patient will visit on the scheduled recall date, calculated based on the care pathways applicable to them. However, in reality the patient's decision to visit may be before, on, or

[5]In this work, we refer to follow-ups or subsequent appointments for a patient as their recalls.

after this scheduled recall; we observed that this behaviour varies across patients and across visits. Hence, we also propose the use of Bayesian inference in our workload prediction model (see Chapter 7).

2. In the process of developing the population-level workload prediction model, we identified shortcomings in current health data models (for example, the LTC status of a patient is not explicit), which is needed to support a population-level workload analysis. We therefore developed a patient information data model.

This work follows a design science research methodology (DSRM). The DSRM address(es) the identified problem(s) in a domain through iterations of a "build-and-evaluate loop" to develop a final design of an artefact with evaluation, feedback, and contributions to the literature occurring at the end of each cycle (Hevner et al., 2004). In this work, the identified problem is the lack of a workload planning tool for a PHC centre. We also present the DSR constructs, the process and the artefact, an adaptable best practice based workload prediction model (ABP-WPM). The iterations to generate this PHC workload management simulation model were primarily driven by feedback from Dr. Hywel Lloyd (hereafter referred to as the "Primary Care Advisor" (PCA)).

1.3 Collaboration and Ethical approval for this work

This work is done in collaboration with the Best Practice Advocacy Centre New Zealand (BPAC), a non-profit organisation that aims to promote best practice for primary health care in New Zealand (see Best Practice Advocacy Centre New Zealand (1997)). The collaborations were in the form of

1. health data from a medical practice;

2. the knowledge base to understand the primary health care domain, and the practical issues at a medical practice.

PCA is a representative of BPAC and is the collaborator of this work. Recently, he changed his role from the Director of Informatics, South Link Health to the Medical Director of Strategy, Primary and Community of Southern District Health Board (DHB). He is a GP himself and also holds an honorary lecturership at the Department of Rural Health and General Practice, University of Otago. Approval for this research using the anonymised patient data provided by BPAC has been given by the University

of Otago Human Ethics Committee (Health) (refer to Appendix A.1 for the approval letter).

A major step in following DSRM is communicating the developed artefact to a broader community (both technology and management oriented audience) including the potential beneficiaries of the study (Hevner et al., 2004; Vaishnavi and Kuechler, 2007). Apart from the feedback from the PCA, feedback was also collected through informal meetings with the care providers of a medical practice at Mosgiel, the members of the Waikato Primary Health Organisation (PHO) including chief medical officers, and members of Southern and Otago DHB including the chief executive officer (CEO), chief information officer (CIO), general manager and others of WellSouth PHO, at various stages of development of this artefact. The artefact developed was also communicated to a research community through two publications (refer to Table 1.1 for the list of publications).

1.4 Organisation

This thesis is structured as follows. Chapter 2 provides an overview of planning care at PHC level and why planning care for a population is important. This chapter also discusses the challenges for an effective PHC workload management and the various workload management systems in health care. This chapter also covers the use of health information technology to support LTC management, with a specific focus on the application of rule-based systems in the health care domain.

Chapter 3 presents our research model and the design science research methodology (DSRM) used to develop our workload prediction model. This chapter covers the holistic view of our research process. We also discuss the cyclic approach followed to develop the workload simulation model according to the DSRM.

The workload prediction model is explained in Chapters 4 and 5. Chapter 4 explains the data pre-processing process and presents the initial patient model that drives the simulation process. This chapter gives more details of the micro-iterations of DSRM that helped to define the rule base for workload prediction. Chapter 5 explains the simulation process and the results of our workload prediction model.

Knowing the workload can help the organisation manage it effectively. Chapter 6 analyses seven what-if scenarios which demonstrate the ability to analyse the impacts of resource bound policies, external factors, changes in practice policies and adopting new models of care and evaluate strategies to address them.

Chapter 7 presents a Bayesian inference approach applied to patient's prescription renewal cycles and their likelihood of visiting early or late for an appointment. This is then applied in the context of predicting patient visits.

Chapter 8 presents current health ontologies and how the shortcomings in current health data models prompted us to develop the patient information data model to support a population-level workload analysis.

Chapter 9 discusses the contribution of this thesis, its limitations and scope for future research.

Figure 1.1 presents the structure of this thesis.

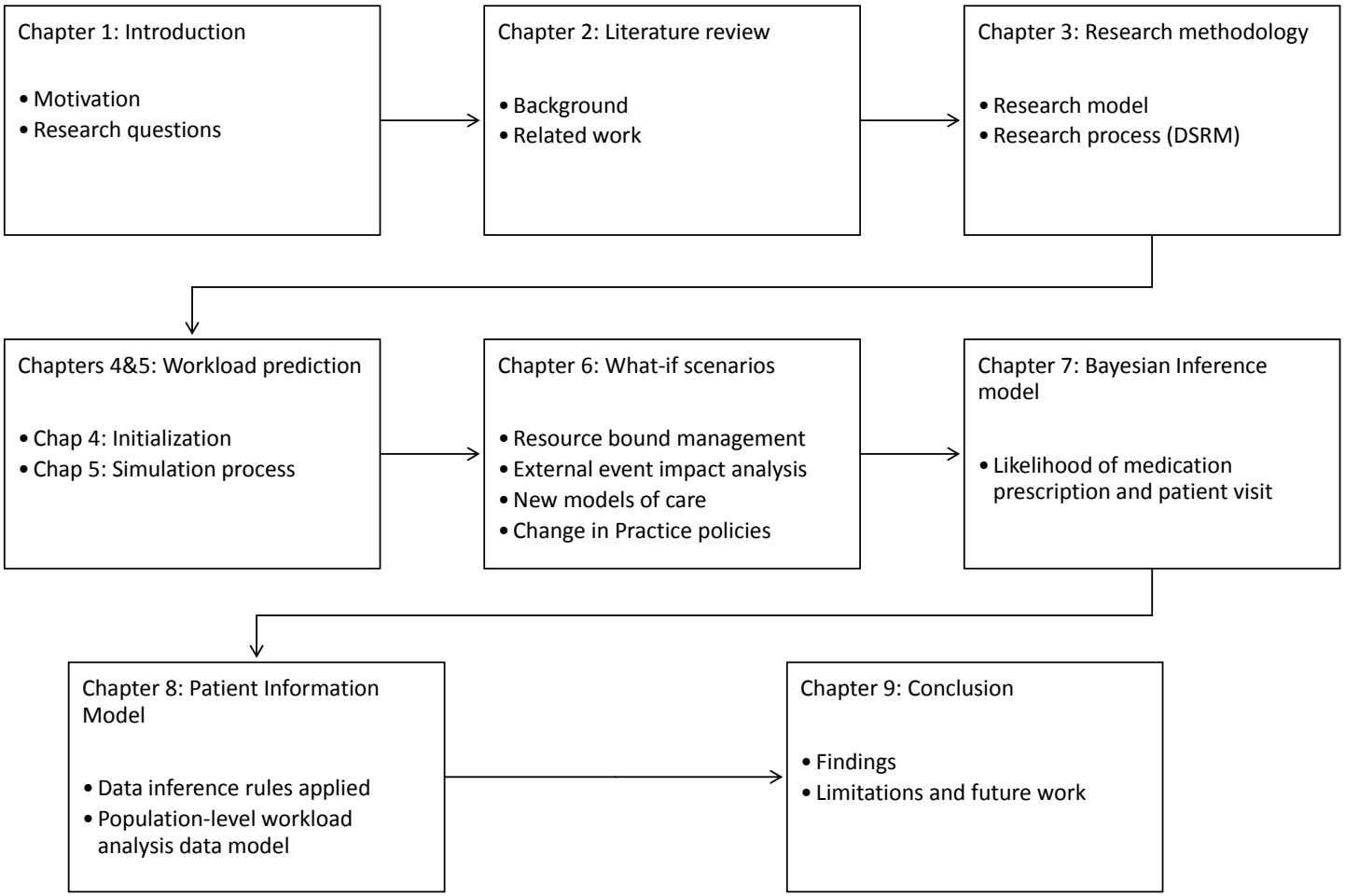


Figure 1.1: Structure of this thesis.

1.4.1 List of publications

The papers published from the work presented in this thesis are summarised in Table 1.1. The table also shows the chapters that each paper is related to.

Table 1.1: Table of refereed publications and corresponding chapters

Devananda, Manjula; Cranefield, Stephen; Winikoff, Michael; and Lloyd, Hywel, (2017). "Workload prediction model of a primary health centre". In Proceedings of the 25th European Conference on Information Systems (ECIS), Guimarães, Portugal, June 5-10, 2017 (pp. 1192-1204). ISBN 978-989-20-7655-3.	This paper presents the rule-based workload simulation model of a PHC centre, initial results of the workload simulation model considering patients with CKD[6] and diabetes only. I developed the workload prediction model and was the primary contributor to the writing. The other authors were my PhD supervisors and the collaborator of this work (Lloyd, Hywel).	Chapters 2, 4 and 5.
Devananda, Manjula; Cranefield, Stephen; Lloyd, Hywel; and Winikoff, Michael, (2017). "Patient Information Model to Support Population-level Workload Analysis". In Proceedings of the 28th Australasian Conference on Information Systems (ACIS), Hobart, Australia, Dec 4-6, 2017.	This paper presents the shortcomings of current PHC data models to enable a population-level workload analysis of a PHC medical practice. It proposes a patient information model developed following Hevner's Three cycle DSRM approach. I developed this patient information model for population-level workload analysis and was the primary contributor to the writing. The other authors were my PhD supervisors and the collaborator of this work (Lloyd, Hywel).	Chapters 2 and 8.

[6]Chronic Kidney Disease

Chapter 2

Background and Related Work

Demand for care is growing dramatically due to an increasing number of people with long-term conditions (LTCs). Long-term conditions are defined as "*any ongoing, long-term or recurring conditions that can have a significant impact on people's lives*" (Ministry of Health NZ, 2017c). Generally, LTCs exist for a prolonged duration (months or years), that tend to eventually lead to associated complications such as developing multi-morbidity in a patient (O'Halloran et al., 2004; Starfield, 2001). Therefore, the care for such LTC patients can span across various levels of the health care system.

A health care system is a multi-tier care delivery system with primary, secondary and tertiary tiers, with the primary health care (PHC) level being the first point of contact with the health care system (Ministry of Health NZ, 2017b; Montague, 2014). While the PHC level provides more "generalist care", the secondary and tertiary care levels provide more specialist and advanced care (WHO, 1978). In other words, these three tiers of care delivery differ in their mode of care delivery, the speciality of their care team, and thus the care itself as well as the cost of the care (Ministry of Health NZ, 2000; WHO, 1978). However, when required, care for LTC management at a primary health care level may involve specialists from the secondary or tertiary levels (Ministry of Health NZ, 2016e). Consequently, for effective and efficient LTC care, the care must be coordinated or sometimes integrated into primary health care planning (Ham, 2010).

As mentioned in the introduction, this work focuses on workload from LTC management needs for patients at the PHC level. Therefore, in this chapter we discuss work related to care at a PHC level, how LTCs are usually managed, the role of IT systems in clinical decision-making regarding care delivery, and what health care workload management strategies are in place or are discussed in the literature.

Sections 2.1 and 2.2 will therefore set the background for this research and the following sections present related work.

2.1 Care at a PHC level

A PHC system is often the first point of contact with the health system for any person (WHO, 1978). In order to access the health care facilities within a PHC centre, the patients initially enrol with a medical practice or a general practitioner (GP) at a medical practice (Montague, 2014). Once enrolled, these patients book appointments with care providers when needed (Johnson, 1997). A health care provider attends to the health care needs of a patient (Johnson, 1997). In other words, a PHC provides a patient-centred, general care to the public, i.e., it aims to meet the individual health care needs of a patient registered with it (Montague, 2014; WHO, 1978). Hence, patient-driven appointments account for the workload at a PHC level.

According to World Health Organisation (WHO, 1978), the primary health care system of a country must implement the Eight Program Elements of PHC, which include immunisations, the prevention, control and appropriate treatment of common diseases and injuries, and health education. Managing the health of a patient may include patient education, timely recalls, repeated medication prescriptions and laboratory tests (Silagy and Weller, 2001). Hence, although a person is enrolled with a GP at a PHC centre, subsequent care delivery may involve other care providers such as nurses and lab technicians. Moreover, in the case when the GP assesses the health care needs of a patient to be beyond the care provided at a medical practice, the patient is referred to secondary or tertiary care (Ministry of Health NZ, 2000; Montague, 2014). Therefore, proactive planning of care delivery at a PHC level is essential.

2.1.1 Terminologies used for planning care

Care delivered at a PHC is usually evidence-based. *Evidence-based care* (EBC) is where a clinician applies both their clinical expertise, gained through experience, and the available external evidence, usually developed through research, experience and technology, to meet a patient's health care needs (Campbell-Scherer, 2010; Silagy and Weller, 2001; Weiner, 2009). This available external evidence is usually published as a *care pathway* which clearly defines the timely interventions, the types of these interventions, and what must be done in order to manage that specific health condition of a patient (Sackett, 1997; Silagy and Weller, 2001).

In order to understand how this evidence-based care is used to plan care for a patient, it is important to understand the concepts of *care pathways*, *care plans* and a *plan-of-care* (also called a *treatment plan*) as used in the health care context. There are several studies, such as those by Calvan et al. (2011); Fox et al. (2006); Johnson (1997); Mathers et al. (2011) and Smith et al. (2012), that use these terms interchangeably. For example, Fox et al. (2006) uses care plans to schedule the activities that should take place to meet individual management needs, while Mathers et al. (2011) refers to a treatment plan to schedule activities to meet the LTC management needs of a patient. Therefore, in this thesis we make a clear distinction between these terms as described below.

The Institute of Medicine (US) Committee on the Future of Primary Care (1996) defined *guidelines* as "*systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances*". Since then, health care agents such as the Guideline Clearing House of the Agency for Healthcare Research and Quality (2014b), the UK National Institute for Health and Clinical Excellence (NICE, 1999) and Scottish Intercollegiate Guidelines Network (SIGN, 2001), review and publish clinical guidelines based on the most recent scientific evidence available. These clinical guidelines, usually represented as flowcharts, are called *care pathways*[1].

While *care pathways* document intervention details, such as when a patient must be recalled, which lab tests should be done and how frequently, to meet the health care needs of a specific health condition of a patient, *care plans* instantiate a care pathway in a patient (Burt et al., 2014; Sackett, 1997). Thus, care pathways focus on standardising the process of care for a specific health condition and care plans are more specific to which path of the care pathway applies to a patient (Best Practice Advocacy Centre New Zealand, 2012; Burt et al., 2014; Calvan et al., 2011; Fox et al., 2006; Government of Canada Health, 2004; Smith et al., 2012).

Generally, a GP addresses every LTC present in a patient (Johnson, 1997). Therefore, in the context of LTC care, in order to plan care that addresses all the LTC management needs of a multi-morbid patient, more than one care pathway has to be considered (Wagner, 1998). For us, this is equivalent to following a step-by-step process, which initially builds a care plan for each LTC in the patient and then merges these care plans to form a single plan-of-care for the patient.

The box below highlights these terminologies and their definitions in the context of

[1]Care pathways are also referred to with many other names, see Johnson (1997) (p. 4)

this thesis.

Care pathway: focuses on a specific LTC's management needs.
Care plan: focuses on a specific LTC's management need in an individual.
Plan-of-care: addresses multi-morbidity management needs in a patient.

2.1.2 Planning LTC care at a PHC level

Wagner (1995) identified that the usual primary care is designed to provide easy access and care for acute cases with an emphasis on patient flow, short appointments and patient-initiated follow-ups. On the other hand managing LTCs in a patient requires more planned, systematic assessments and attention to treatment guidelines. Most often, intervention by various care providers is inevitable to manage multi-morbidity in a patient (Reeves et al., 2014). It is clear that when multiple such roles intervene, they must work collaboratively and coordinate to plan care and, would need to refer to the same plan-of-care developed for a patient (DHS Primary Health Branch Victoria, 2008; Harris and Zwar, 2007). Hence, an evidence-based care planning process is needed to develop a shared *plan-of-care* for an LTC patient (Amir et al., 2015; Johnson, 1997; Weiner, 2009). Thus, care planning is a process that may involve multi-disciplinary roles, to plan care for a patient, and a plan-of-care is an outcome of that process.

A plan-of-care developed following a comprehensive care planning process plays a vital role in the quality of care delivered (Young et al., 2017). Studies (Amir et al., 2015; Reeves et al., 2014; Utley and Worthington, 2012; Weiss, 1998; Woolf et al., 1999) show that following a care planning process for LTC management has many advantages, such as:

- multi-morbidity in an LTC patient is addressed through a single plan-of-care for the patient; therefore, a care planning process will be patient-centred rather than treating LTCs in isolation.

- it involves multidisciplinary roles to plan and deliver care, and thus avoids fragmented care and ensures continuity of care for an LTC patient;

- adherence to a plan-of-care avoids clinical errors such as repeated lab tests and prescribing interacting medications for an LTC patient. Planning of care will consider all LTCs present in a patient, which in turn, will have a view of all the care actions needed; for example, a list of all the lab tests required, and thus a care provider can easily avoid repeated lab tests.

- it clearly articulates the role of each care provider of the patient. This transparency invokes coordinated, integrated care as well as motivates an LTC patient to manage health better (Babiker et al., 2014), which in turn leads to better health at a lower cost.

Wagner (1995) claims that rather than planning care for an individual, planning care for a population of patients would benefit chronic patients. Later in 1998, he proposed a theoretical framework, the chronic care model (CCM) that clearly specifies the elements of change required to deliver effective and efficient chronic care. As described below, Bodenheimer et al. (2002a) then enhanced the CCM further to add patient safety, care management and case management to the CCM model. Therefore, since it is widely accepted (Table 2.1 lists a few case studies) to improve quality of chronic care delivered, in the following subsection I will describe the components of CCM and which aspects of CCM are incorporated in the work presented in this thesis.

Table 2.1: Some case studies that incorporated CCM components

Name	**CCM components implemented**	**Chronic illness considered**
Premier Health Partners (US) (Premier Health, 2018)	Decision support, physician performance feedback	Diabetes
HealthPartners Medical Group (US) (HealthPartners, 2018)	Disease registry, case management, primary care teams	Diabetes
Clinica Campesina (US) (Clinica, 2018)	Self-management, disease registry, reminder systems and primary care teams	Diabetes, cardiovascular disease, asthma, and hypertension
Kaiser-Permanente Northern California (US) Pines et al. (2015)	Primary care team, disease specific care management, asthma registry, self-management for asthma patients, clinical information systems	Diabetes, coronary artery disease, hyperlipidemia, asthma and congestive heart failure

Chains of Care (Sweden) (Åhgren, 2003)	Primary care teams, care-management, delivery system design	Patients with similar conditions were grouped together
Disease Management Programs (Germany) (Institute for Quality and Efficiency in Health Care, 2016)	Evidence-based guidelines, clinical information systems, health care organisation, self-management, population-based prevention oriented	Diabetes
Managed Clinical Networks (UK) (Skipper, 2010)	Coordinated care, link between various levels of health care system, health care organisation, delivery system	Cardiac services, neonatal care and cancer.

2.1.3 The Chronic Care Model (CCM)

In 1995, Wagner identified that complex needs due to multi-morbidity requires more planned, systematic assessments and attention to treatment guidelines and so proposed the chronic care model (CCM). In his CCM, he highlights six components required to improve care for chronic patients.

As given in Table 2.2, the components emphasise the role of community resources such as exercise groups and patient education classes. CCM also mentions how to reorient a health care organisation, specifically the link between the care provider organisation and the insurance providers and other stakeholders. He portrays in his CCM that self-management of LTCs plays an important role in chronic care.

In order to self-manage LTCs in patients, they must be educated and trained to exercise, diet and use monitoring tools such as blood pressure cuffs and glucometers. Wagner claims that creating care teams with a clear division of roles and responsibilities, including training non-physician roles to support patient care, can improve care delivery. He also asserts that evidence-based care must be integrated into the daily care delivery process to standardise and systematise chronic care.

Finally, the CCM emphasises the use of computerised information systems to augment care delivery. Wagner claims that computerised information can be used effectively to generate timely reminders for the follow-ups, to maintain disease registries for population-based information and, also as a feedback system for the physicians to

Table 2.2: Components of the CCM with their purpose and examples.

CCM components	**Purpose of the CCM component**	**Examples**
Community resources and policies	To associate care provider organisations with community-based resources	Exercise programs, senior centres, self-help groups
Health care Organisation	Structure the values and goals and its relationship with insurers to prioritise chronic care	Reimbursement environment, purchasers and insurers
Self-management Support	Involves collaboratively helping patients and their families to acquire skills and confidence to manage their conditions, provide self-management tools such as blood pressure cuffs and glucometers	Diet, exercise, self-monitoring, medication use
Delivery System	Creating practice teams with clear roles and responsibilities.	Non-physician personnel are trained to arrange routine periodic tasks such as foot examinations.
Clinical Information Systems	Plays three important roles namely as reminder systems to help comply with practice guidelines, as feedback system to physicians and as registries for planning individual patient care and conducting population-based care.	Registries lists all patients with a particular chronic condition
Decision Support	Evidence-based clinical practice guidelines are integrated into daily practice	Guidelines are reinforced for practice teams.

reflect on how their patients are managing their LTCs, which in turn, can help improve effectiveness and efficiency of care delivered.

Later, Bodenheimer et al. (2002b) enhanced the CCM with an emphasis on patient safety, case management and care management. It is thereafter referred to as the *enhanced CCM model*. Ever since the enhanced CCM was introduced, it is seen as a guide to design and deliver effective chronic care. It advocates organising principles that help identify the changes required to improve care coordination and for the care delivered to be evidence-based, population-based and patient-centred for chronic patients (Bodenheimer et al., 2002a; Glasgow et al., 2001).

2.2 Adopting CCM components at a PHC level

Although, the case studies in Table 2.1 implement only some aspects of the CCM in their care delivery, they demonstrate that adopting one or more components of CCM improved quality of care. Moreover, based on Wagner (1995) and Kane et al. (2005), Ham (2010) describes ten characteristics of a high performing chronic care system which emphasise care teams, population-based care, emphasis on evidence-based care, integrated care and coordinated care for chronic patients. Therefore, we chose to implement some components of CCM components in our workload prediction model. Table 2.3 summarises features of our workload management model and showcases the various components of CCM being considered through those features.

In our workload management model, we use the best practice guidelines for managing LTCs as a tool to decide the frequency of recalls of the patient. We then aggregate these predicted recalls on a weekly basis for a population-level workload analysis. In order to identify which care pathways apply to these patients, to predict the upcoming workload from these LTC patients and, to analyse the predicted workload, we use rule-based systems. It also thus makes our proposed model align with the characteristics of a high performing chronic care model.

The rest of this section, therefore, discusses these aspects, namely, evidence-based, patient-centred, population-based care, team-based care, and the use of IT in the health care domain. In Section 2.2.4 where we present work related to team-based care, we will also discuss work related to coordination and integrated care.

Table 2.3: Summary of key features of our workload management model that follow from the CCM.

Features of our model	**Components of the CCM considered in our model**
Care pathways, care plans and plans-of-care	Evidence based care
Plans-of-care, recalls of a patient	Patient-centred care
Population-level workload	Population-based approach
Best practice guidelines	Evidence-based, population-based approach
Shifting patients among various roles	Team-based care
Stratify patients based on severity	Can help to coordinate and integrate care
Rule-based system	Use of computerised systems for decision-making

2.2.1 Evidence-based care (EBC)

I briefly introduced evidence-based care in Section 2.1.1 to distinguish various terminologies used in the literature and our choice of terminology. In this section, I discuss evidence-based care as more than just the terminologies used for planning care.

During an appointment, a clinician assesses the health care needs of the patient (Silagy and Weller, 2001). Most often, multi-morbidity requires the clinician to consider, simultaneously, the evidence for treatment (i.e., the care pathways) for every LTC in the patient (Bodenheimer et al., 2005, 2002a,b). When more than one care pathway applies, selecting which interventions to prioritize can be difficult (Calvan et al., 2011). Furthermore, it can be even more difficult to determine the effectiveness of the care delivered (Fetherstonhaugh et al., 2013). The expertise of a clinician aids in making such complex health care decisions (De Bleser et al., 2006). This is specifically the case with multi-morbidity, where the potential for medication interactions, unwanted or unanticipated effects or contraindications to interventions are greatly increased (Sackett, 1997).

Moreover, self-management is an essential part of LTC management in a patient (Wagner, 1998). The patient and the clinician have to come to a consensus on the treatment plan, which may include change in the patient's life style (Burt et al., 2014). This treatment plan should incorporate patient preferences as well (Silagy and Weller,

2001). Therefore, in order to apply existing evidence to individual patient situations, the most valid approach is to embrace evidence-based patient choice (Johnson, 1997; Sackett, 1997; Silagy and Weller, 2001).

Evidence-based care blends a clinician's experience with available external evidence to decide the care required for a patient's well-being and health, taking into consideration the patient preferences (Campbell-Scherer, 2010; Fetherstonhaugh et al., 2013; Sackett, 1997; Silagy and Weller, 2001; Weiner, 2009). It is clear that with evidence-based care, use of external evidence can eliminate the practice risks that may occur otherwise. And, with clinical expertise, care can be planned even when excellent external evidence may not be applicable to a patient (Sackett, 1997).

Sackett (1997) claims that evidence-based care can track down which evidence best applies in a situation, and integrate it with clinical expertise and patient preference, and can evaluate the performance. However, Fetherstonhaugh et al. (2013) point out that evidence about effectiveness of an intervention, i.e., that it works in a controlled population, does not answer the question about applicability or feasibility in a specific client-clinician context. So, generally, evidence-based care is applied in conjunction with patient-centred care.

2.2.2 Patient-centred care (PCC)

The Institute of Medicine (US) Committee on Quality of Health Care in America (2001) defines patient-centered care as: *"providing care that is respectful of, and responsive to, individual patient preferences, needs and values, and ensuring that patient values guide all clinical decisions"* (p. 6). Robinson et al. (2008) identifies the fundamental characteristics of patient-centered care (PCC) as individualised care involving the patient in the care planning process. Thus, PCC considers the patient as central to care provision, which requires the patient to be involved in their care and to have the knowledge, skills, and motivation to do so (Bilello et al., 2018). Consequently, in PCC, the doctor-patient interaction will have a shared goal, written plan-of-care and regular follow-up (Bauman et al., 2003).

Considering patient preferences in health care decisions (evidence-based care) is different from involving patients in the care process (patient-centred care). For instance, a patient may not agree to exercise every day (a patient preference) but may agree to self monitor blood pressure (a patient involvement in care). Therefore, PCC emphasises that the treatment plan should be organised around whole-person goals (e.g., school readiness) rather than organ-system goals (e.g., brain, lung) (Burt et al., 2014).

At a PHC level, there are many LTC patients. Hence, planning care for a group of patients at once is useful (Bodenheimer et al., 2002a). This approach to consider care for a group of people is usually referred to as a *population-based* approach (Kindig and Stoddart, 2003). Wagner (1995) and Ham (2010) in their studies emphasise that patient-centred care along with population-based care can improve effectiveness and efficiency of care for LTC patients.

2.2.3 Population-based care (PBC)

Population-based care (PBC) considers care for a group of patients who share a common trait (Wagner, 1995). This common trait could be sharing the same GP, being diagnosed with the same LTC, or following a similar plan-of-care (WHO, 2005, p. 45). For planning care, the needs of a population (or a cohort) is identified to plan, organise and deliver care accordingly (Ibrahim et al., 2001; Wagner, 1995). In other words, PBC planning develops a plan-of-care that ensures systematised care for a group of LTC patients at a medical practice (Institute of Medicine (US) Committee, 2002). Therefore, population-based care involves

1. identifying (defining) the population; i.e., the health care needs that can be addressed for a group of patients.
2. involving a multidisciplinary team, and
3. use of IT systems to promote, provide and evaluate population-based care (Grundy and Hodach, 2016; Weiss, 1998).

Although, PBC can be provided by multiple physicians, usually a (chosen) GP serves as the central figure for delivering population-based health care to the entire community (Weiss, 1998). Ibrahim et al. (2001) list five characteristics of a population-based approach for care: "a community perspective, a clinical epidemiology perspective using population-based data, evidence-based practice, an emphasis on outcomes, and an emphasis on prevention". Furthermore, a high performing chronic care system is also characterised by population-based care and adherence to best practice for the group of patients (Ham, 2010; Wagner, 1995).

With a good understanding of the common health care needs of a sub-population of patients, care planning also involves coordinating the intervention of various health care providers and in turn meeting the characteristics of PBC (Kindig and Stoddart, 2003; Payne et al., 1995; Public Health Agency of Canada, 2001). For example, following PBC

for diabetic patients, a medical practice can plan a week for the annual diabetic checks and screening for other associated complexities, which may involve availability of a dietitian or an exercise coach as needed (Ministry of Health NZ, 2016a). Consequently, medical practices will perform more efficiently with a population-based care model for their LTC patients.

While patient-centred care considers the patient as a whole, population-based care considers commonalities in health care needs of groups of patients (Kane et al., 2005). Planning care for a patient considers multiple care pathways applicable to the patient, while a population-based care would consider a care pathway applicable to a group of patients and the common needs of these group of patients (Wagner, 1995). In the context of LTCs, a chronic condition can be related to another in different ways: they are independent or one is an associated complication of the other (Ash et al., 2000).

2.2.4 Team-based care (TBC)

According to the Ministry of Health NZ (2016b, 2017c), within an effective PHC system for LTC management, GPs and other care providers work collaboratively to identify the severity of an LTC, address associated complications and take measures to avoid the occurrence of other LTCs in a patient. It is evident that involvement of multidisciplinary roles makes LTC management better (Stellefson, 2013). Apart from those by Wagner (1998) and Bodenheimer et al. (2002a), a few other studies emphasise the requirement of creating care teams to work closely with GPs to offer LTC management support, including specialist care and self-management advice (Ash et al., 2000; Babiker et al., 2014; Dale, 2015; Mays, 2013; Ministry of Health NZ, 2000).

The common strategy in team-based approaches (e.g., Patient-Centered Home (Robinson et al., 2008), Health Care Home Model (Grant and Greene, 2012) and Family Medicine Group Home Model (Breton et al., 2011)), is that the primary GP of a patient continues to offer health services, but as a member of a team of care providers who work together under "one roof" (Babiker et al., 2014). This team is willing to communicate and share the plans-of-care for their patients, and thus share accountability for their patients within the team (Babiker et al., 2014; Mickan and Rodger, 2005). The patient, having been made aware of this shared plan-of-care and its importance, is ready to follow up, in accordance with their timely recall, with any member of the care team (Payne et al., 1995). Thus, having a care team ensures continuity of care and improves quality of care (Aysola et al., 2015; Davy et al., 2015; Ham, 2010; Stokes et al., 2017). Although team-based care involves multiple care provider roles, much

emphasis is on coordinating and integrating care for their LTC patients (Davy et al., 2015; Mickan and Rodger, 2005).

Care coordination has been mentioned in various studies (Ash et al., 2000; Babiker et al., 2014; Stellefson, 2013): however, there is no consensus on its definition and scope of application (Agency for Healthcare Research and Quality, 2014a). For instance, while some of these studies highlight the need for communication among various care providers, patients, and patient families, as needed to plan care for the patient (Åhgren, 2003; Babiker et al., 2014), a few others use care coordination to bring forth the idea of *shared decision making* to decide the interventions required for the patient (Baud, 2003; Bauman et al., 2003; Harris and Zwar, 2007). For us, we consider care coordination to be when more than one care provider shares the responsibility of care for a patient or a group of patients amongst themselves.

While care coordination focuses on what is required to be agreed on for an effective plan-of-care for a patient, integrated care focuses on the roles, and the scope of those roles that are involved (Busetto et al., 2017; Young et al., 2017). Specifically for LTC patients, as the complexities of health care needs increase, multiple roles spanning across various health care levels need to be integrated into the care planning process (Young et al., 2017). For instance, for a diabetic population, in order to address the associated complexities, care planning needs to involve roles such as ophthalmologists (for diabetic eye tests) and podiatrists (for foot checks) (Best Practice Advocacy Centre New Zealand, 2012). These roles may be in the secondary or tertiary health care levels (Ministry of Health NZ, 2016a).

In summary, team-based care with coordinated and integrated involvement of various roles can help enhance the efficiency and effectiveness of PHC.

2.2.5 Use of IT in health care

Yet another characteristic of a high performing health care system is utilisation of information technology (IT) to the fullest in improving chronic care (Ham, 2010; Wagner, 1998). The use of IT in health care spans from developing formal concept models (ontologies) (NHS Digital, 2017a; openEHR, 2016) to using sophisticated measures like deep learning (Archenaa and Anita, 2015; Kim et al., 2018; Krittanawong et al., 2018) to utilise available data to make informed health care decisions.

The use of ontologies[2] in medicine is mainly focussed on the representation and

[2]Ontologies mean concepts and the relationship between these concepts

(re-)organisation of medical terminologies. Physicians developed their own specialized languages and encodings such as Read codes (NHS Digital, 2017a), SNOMED-CT codes (NHS Digital, 2017b) and ICD-10 (WHO, 2004) to help them store and communicate general medical knowledge and patient-related information efficiently. Such terminologies are characterized by a significant amount of implicit knowledge. Medical information systems need to be able to communicate complex and detailed medical concepts (possibly expressed in different languages) unambiguously. Health data models such as HL7 (HL7, 2016) and openEHR (openEHR, 2016) were developed to achieve this requirement. Moreover, in order to achieve specific requirements of various studies, e.g., ChronoMedIt (Mabotuwana and Warren, 2010), they define study-specific ontologies in health care domain. Section 8.1 presents a detailed discussion on the current health ontologies (used or referred to in this work).

Deep learning techniques are recently gaining attention due to their applications in data-driven decisions. Using deep learning is known to improve efficiency (through multi-layer processing with less time and better accuracy performance) over other competing methods such as logic regression and decision trees (Kim et al., 2018). It is also widely used in contexts such as to predict the health deterioration of a specific condition e.g., hypertension in a patient (Krittanawong et al., 2018), or to perform a real-time analysis of patient health data and avoid emergency situations (Archenaa and Anita, 2015). However, we focus on computer based clinical decision support systems where knowledge rich clinical guidelines play an important role in decisions related to one's care delivery. In contrast deep learning uses historic and current health data to extract features to represent its knowledge base.

Advancement of IT in health care has led to acceptance of computer-based clinical decision support systems (CDSSs). These CDSSs are built on formalised clinical guidelines. Shiffman (1997) justifies the use of rule-based models in formalising clinical guidelines.

Rule based systems are built on fixed "When-If-Then" formulae, and draw conclusions based on all the facts in their working memory (Kuo and Fuh, 2009). For example, a diagnosis of various stages of diabetes follows:

- if HbA1C below 6.0%, then normal (no diabetes)
- if HbA1C between 6.0% to 6.4%, then pre-diabetes
- if HbA1C 6.5% or above, then diabetes.

Therefore, knowledge-based clinical decision support systems use an "If-Then" rule base applied onto the patient data to propose care actions for the patients (Alther and Reddy, 2015). In addition to Shiffman (1997), Minutolo et al. (2017) also highlight the benefits of using a rule-based system in medical settings, which include:

1. flexibility to choose different rules based on the facts expressing the current context, and

2. dynamic handling of changes through rules being re-activated based on changes to facts during the execution of rules.

Calcaterra et al. (2018); Fernandez-Millan et al. (2015); Lin et al. (2018) and Semenov et al. (2018) underline the strength of rule-base systems to provide physicians with real-time guidance for health care decisions with an emphasis on the role of clinical guidelines as rules to achieve a patient's health care encounter outcomes.

Table 2.4 summarises the studies we have referred to under each of these CCM components. Although we discussed these components one-by-one, most studies that implement CCM components implement more than one component. Although they might refer to more than one component in their study, we list them under specific CCM components based on the emphasis they give to that CCM component in their work.

Table 2.4: CCM components emphasised in other studies referred to in this section

Work referred to	**EBC**[3]	**PCC**[4]	**PBC**[5]	**TBC**[6]	**Use of IT**
Bodenheimer et al. (2005); Calvan et al. (2011); Campbell-Scherer (2010); De Bleser et al. (2006); Fetherstonhaugh et al. (2013); Johnson (1997); Sackett (1997); Silagy and Weller (2001)	X				
Bauman et al. (2003); Bilello et al. (2018); Institute of Medicine (US) Committee on Quality of Health Care in America (2001)		X			

[3]Evidence Based Care
[4]Patient Centred Care
[5]Population Based Care
[6]Team Based Care

Grundy and Hodach (2016); Ibrahim et al. (2001); Institute of Medicine (US) Committee (2002); Kane et al. (2005); Kindig and Stoddart (2003); Public Health Agency of Canada (2001); Weiss (1998); WHO (2005)			X		
Åhgren (2003); Babiker et al. (2014); Breton et al. (2011); Busetto et al. (2017); Dale (2015); Davy et al. (2015); Grant and Greene (2012); Mays (2013); Mickan and Rodger (2005); Ministry of Health NZ (2000, 2017c); Payne et al. (1995); Stellefson (2013); Stokes et al. (2017); Young et al. (2017)				X	
Alther and Reddy (2015); Archenaa and Anita (2015); Calcaterra et al. (2018); Chute (2000); Fernandez-Millan et al. (2015); Goertzel (1969); Kim et al. (2018); Kuo and Fuh (2009); Lin et al. (2018); Ministry of Health NZ (2017j); Minutolo et al. (2017); NHS Digital (2017a); ONC (2016); openEHR (2016); Semenov et al. (2018); Shiffman (1997); Spackman et al. (1997); Tsiknakis et al. (2002); WHO (2016a)					X
Burt et al. (2014)	X	X			
Best Practice Advocacy Centre New Zealand (2012)	X			X	X
Aysola et al. (2015); Robinson et al. (2008)		X		X	
Wagner (1995)		X	X		
Ash et al. (2000); Ministry of Health NZ (2016a)			X	X	
Bodenheimer et al. (2002a,b); Ham (2010); Wagner (1998)	X	X	X	X	X

2.3 Related work

The main aim of our model is to predict (or forecast) the upcoming workload (activities to be scheduled) from LTC patients. In order to schedule the activities, it is important to know what these activities are and who can perform these activities (Decouttere and Vandaele, 2014).

Given, that primary health care is generalist care, which means it covers health care needs related to short acute consultations, immunisations, LTC management needs and referrals to secondary and tertiary care levels (WHO, 1978), there is a huge body of

literature in this domain. Therefore, in order to narrow down the literature, we scoped our search for related work specifically to research that helped us identify what the workload variables are and how the various services related to care are delivered. We also discuss work related to simulation in health care.

One of the benefits of having a tool that can anticipate future workload is that it can help the organisation equip itself with the right mix of roles to manage the upcoming workload. Therefore, we also looked at different workload forecasting, analysis and management tools or techniques. Finally, I compare each work discussed in this section to the focus of our work.

2.3.1 Workload at a PHC centre

It is very important for practices to understand patient demand and manage their workloads well (Anonymous, 2006b). In the context of LTC management, much emphasise is placed on educating patients, involving patients in their care planning process, and coordinating and integrating care within the care team (Babiker et al., 2014; Bodenheimer et al., 2002a; Wagner, 1998). However, variance in practice, shortage in workforce, short appointments, fixed policies around roles and responsibilities, condition based funding structure, and excessive use of IT such as use of health apps by patients hinders adopting an efficient and effective model of care for their LTC patients (Harris and Zwar, 2007; O'Leary et al., 2013; Reeves et al., 2014; Terry, 2017; Trindade and Pires de, 2013).

In a PHC context, the volume of care, including the number of visits, reflects the PHC's effectiveness and efficiency (Anonymous, 2006b; Heroman et al., 2012). One of the ways to anticipate this volume of care is to analyse how the visits of patients are planned (Jordan et al., 2003; Ministry of Health NZ, 2017k).

Hall (2012), in an introduction to a Handbook of Healthcare System Scheduling, gives an overview of issues and options in scheduling healthcare resources. Table 2.5 shows a comparison of a few questions asked by Hall (2012) with our research questions. He notes that health conditions emerge randomly due to various reasons such as an injury or illness, and the patients do not always show-up promptly. In this way, he projects that uncertainty in patient visits should be addressed to achieve best-planned workload management schedules. In contrast, we model workload at the population level, and so do not consider changes in health conditions, but address uncertainty in the timing of patient visits. Moreover, his questions focus on the emergency department of a hospital and also have more of a patient perspective rather than a population

perspective.

Table 2.5: Comparison of questions addressed in Hall (2012) and our research questions.

Hall (2012)	**Our study**
How many patients can we expect to present for care in an emergency department on a given day of week, time of day and time of year?	How many patients are expected to visit the practice during a given period of the year? (We use a weekly basis to present the results).
What is the projected future need to care for a patient of a particular age, weight and blood glucose level, who has been diagnosed for type 2 diabetes?	What is the future number of appointments from patients with low severity of LTCs in them? (Addressed in one of our what-if scenarios).
How will the demand for appointments depend on the prevalence of influenza, given the time of year and cases seen to date?	What is the impact of (a 3-week long) flu outbreak on the predicted workload? (Addressed in our what-if scenarios).
How likely is it that a particular patient will be a no-show for a scheduled appointment, made a set number of weeks in advance?	How likely is a patient to visit the practice knowing when their medication would run out? (We analyse this through a Bayesian model)

Fox et al. (2006) present the CREDO framework for cancer care to "evaluate the use of decision support and workflow services at many points in the *patient journey* ...". A patient journey comprises an *initial encounter*, *work-up*, followed by delivery of treatment, and long term *follow-up*. The services advocated in this CREDO framework include providing personalised schedules. These personalised schedules are embedded within a patient's care plan[7]. As described in Section 2.1.1, care plans in the CREDO framework is equivalent to our plan-of-care. The CREDO framework focuses on how the services are delivered rather than anticipating what services will be required (the aim of our work).

Mathers et al. (2011) characterise care planning as an "example of putting self management support into practice, in a systematic way, as part of routine care for people with LTCs" (p. 9). They highlight that care planning can help collect and

[7]For Fox et al. (2006) care plans are plans of future activities, specific to a patient's problem(s), treatment and goals, which are signed and time-stamped (p. 838).

aggregate information such as the services needed and who needs those services, and thus can determine the unmet health care needs of their population. Just as they aggregate data to decide to integrate self-management into their care planning process and monitor outcomes of the care planning process, we aim to aggregate the patient's health care needs to determine the workforce changes needed or the alternatives in practice policies that could help meet the upcoming workload.

Brown et al. (2018) analyse how care planning is applied in a primary care context, how multi-morbid patients benefit from care planning and in what contexts care planning works best for the patients. Thus, their study focuses on the patient perspectives. In our case, we aim to use care planning (planning patient visits of LTC patients including multi-morbid patients) in order to estimate the upcoming workload as a number of appointments required per week. Just as a patient, being aware of the impacts of lab results and what needs to be discussed during consultation, can be actively participating in the care planning process, we argue that knowing the upcoming workload can help care providers to be more proactive in planning care for their LTC patients.

There are studies that use patient data and historic visit information in mathematical or statistical tools to forecast demand for care. For instance, Green (2013); Green and Savin (2008) and Utley and Worthington (2012) use queueing models, while Abdel-Aal and Mangoud (1998); Potts et al. (2011) and Murray and Berwick (2003) use statistical models including Bayesian models. However, little is discussed about adherence to guidelines and patient-centredness. On the other hand, we use Bayesian Inference along with adherence to clinical guidelines to predict the upcoming workload (refer to Chapter 7 for our Bayesian inference model).

In summary, although previous work did not address evidence-based care or patient-centred care in their studies, attempts to forecast demand for care using mathematical and statistical models show that the capacity planning issue existed decades ago. Some other studies show that treatment plans can be used to schedule activities to manage individual health care needs (Amir et al., 2015; Burt et al., 2014; Fox et al., 2009; Mathers et al., 2011). These activities, in turn, help to understand the volume of required and delivered care. Moreover, the implementation of a care pathway has reduced the variability within the clinical practice, reduced health care costs and improved patient outcomes (Best Practice Advocacy Centre New Zealand, 2012). Thus, our study using plans-of-care, developed from care plans using care-pathways (see Section 2.1.1) to anticipate the upcoming number of visits from LTC patients addresses the requirements

of an effective and efficient chronic care model.

2.3.2 Models of Care

Yet another aspect of care delivery is to address how the health care services are delivered to those who require those services. The term "Models of Care", though definitions vary, is usually a general term referring to how the health care services are provided to a patient or a population of patients in a specified care setting (Davy et al., 2015; Jackson et al., 2013; Stokes et al., 2017). Therefore, based on the health care setting and whether the services are provided for a patient or a group of patients, the model of care applied varies. There are also models of care that emphasise care teams. It has to be noted that these models of care, most often, do not explicitly refer to whether they can be applied to a single patient context or not.

While Wagner's Chronic Care Model (CCM) is a generic model widely accepted for chronic care, there are other models of care that emerged to address various aspects of care. For instance, the Chain of Care was developed in Sweden to address the problem due to fragmented care (Åhgren, 2003) and the Managed Clinical Networks (MCNs) was developed in Scotland to bring together health professions from primary, secondary and tertiary care to work collaboratively to provide high quality clinically effective services (Skipper, 2010). For an effective Chain of Care, the values of the health care provider, involvement of patients and the agreed activity plans are important. Thus, it emphasises patients as contributors to their care rather than as receivers. On the other hand, MCNs consider the expertise of health professionals and coordination among them as important factors. While the role of patients in a Chain of Care is explicit, MCNs assume that their patients benefit from ease of access and continuity of care (Åhgren, 2003; Skipper, 2010).

The earliest population-based model of care was the Kaiser-Permanente (KP) Risk Stratification Model (Pines et al., 2015). It was the first model of care to consider grouping patients based on their health care needs. According to the KP stratification model, the chronic disease population is organised into three categories: people requiring standard care and support to self manage; people needing regular contact with a multidisciplinary team to ensure effective management of their disease; and people requiring more intensive support, perhaps from a specialist case manager, often when they are coping with the complications of co-morbidities (WHO, 2016b). According to Pines et al. (2015), the KP model has three distinct inter-related business entities, namely a health plan, medical groups of physicians and a hospital system. Figure 2.1

shows that the stratified population is considered for the health plan (Kaiser Foundation Health Plan, or KFHP), while medical groups and hospitals are linked to the members of the care funding team and share a central view of the KFHP. Therefore, the KP model is focused on chronic care within a multi-speciality practice rather than primary and secondary care level. However, there is some work that partially incorporated KP into their primary health care delivery context (e.g., ICARE4EU (van Ginneken and Rijken, 2016)).

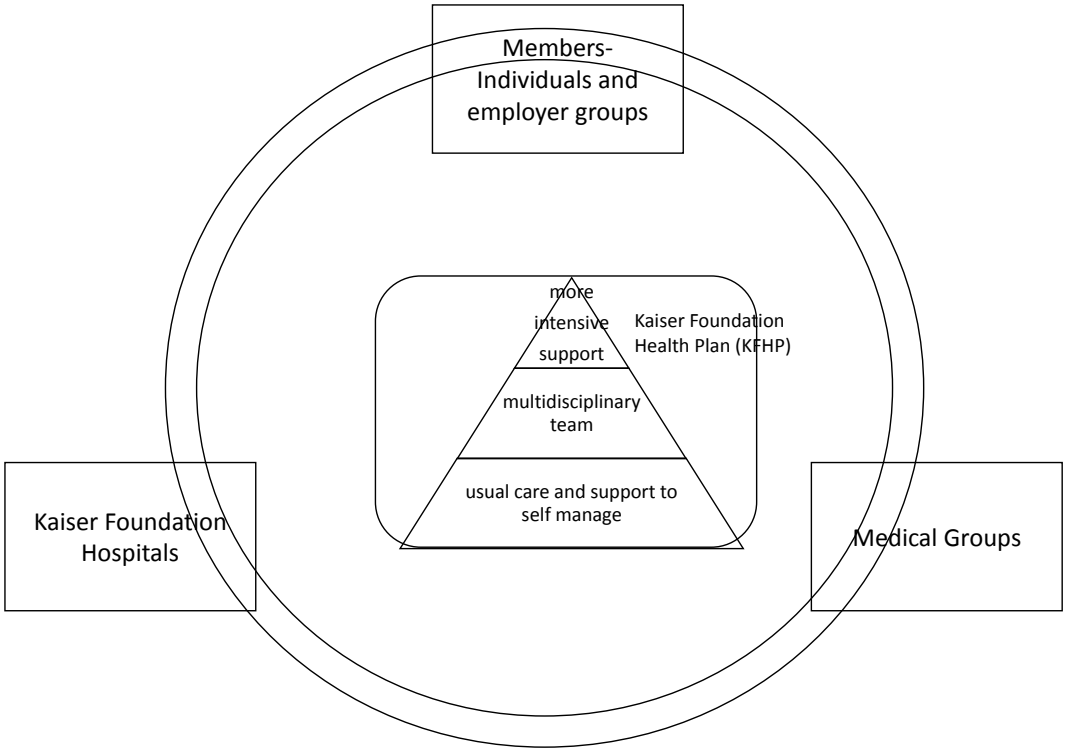


Figure 2.1: KP model from WHO (2016b) incorporated with the business entities from Pines et al. (2015).

Similar to the KP stratification model, Struijs et al. (2015) developed the Care Continuum Alliance (CCA) model, a model for population health at the primary care level. It introduces a best-practice framework that depicts six steps of shifting patient-centred care to population-based care (see Figure 2.2). It starts by identifying the population, then performs health assessment on the identified population. The third step stratifies the population based on the assessment in Step 2. Step 4, patient-centred interventions, covers interventions across the whole spectrum; from public health interventions to stimulating healthy lifestyle till palliative care interventions to provide the best possible quality of life for people approaching the end of life. Step 5 then

evaluates the effects of these interventions. Step 6 is the quality improvement process which is a feedback loop that helps to refine the processes that identify the population, their commonalities and care activities required to meet their LTC management needs and thus reflect on the patient and the patient-clinician interaction.

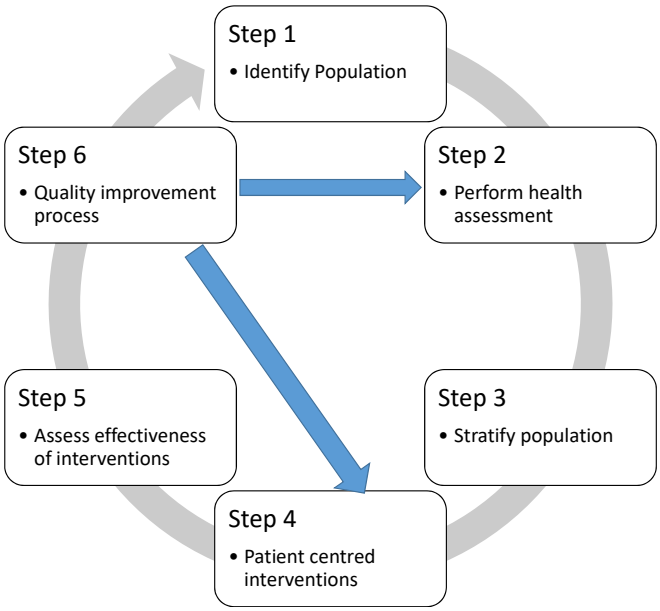


Figure 2.2: CCA framework redrawn based on (Struijs et al., 2015, p. 525).

The House of Care (Taylor, 2015) is yet another population-based model of care, specifically for LTC patients. It considers the local resources available, including the non-clinical staff for planning care to support LTC patients in a community. This model of care comprises two components, namely *execution* and *realisation*. The execution part deals with the practical issues of integrating the new model of care, while the realisation part reduces the risks of change due to the incorporation of a new model of care. However, due to the ever-changing nature of the demands of care and almost fixed health care policies of an organisation, the flexibility to adopt this model of care in different PHC contexts is unknown.

These population-based models of care are based on the idea that stratifying the patient population based on their health care needs will help to equip local resources, and train the non-clinical staff to support planning care for their patients. Moreover, having a best-practice framework will lead to better coordination and integration of health care professionals into the planning process.

Patient-Centred Medical Home (PCMH), developed from Medical Care Home Models, is a physician-directed practice that provides accessible, continuous, comprehensive,

coordinated and delivered care within the context of family and community (Grant and Greene, 2012; Reid et al., 2009; WHO, 2016b). It features personalised care in which the patients know who is responsible for their health and the providers know the patients they are responsible for (Reid et al., 2009). PCMH is often seen as a model of care for coordinating care across various levels of health care. It involves multi-disciplinary care teams as well as, if required, primary physician buying care from a specialist team (Jackson et al., 2013). This integration of care invokes a shared responsibility for the health of their patients. Reid et al. (2009) in their study, found PCMH improves the patient experience, and quality of care with less staff burnout.

New Zealand's health model, built on the Ottawa charter (WHO, 1986) for health promotion, highlights that health promotion actions and planning must consider the local needs of people rather than following "one size fits all" strategy. Therefore, in order to incorporate Māori cultural aspects, New Zealand has three additional models of health (Ministry of Health NZ, 2017d). The first one addresses the health care needs from four different perspectives, namely physical, spiritual, family and mental health (Ministry of Health NZ, 2017f). The second model is represented as an octopus with the head representing family, eyes the individual well-being and the eight tentacles depicting various dimensions of health for well-being. These eight dimensions represent "spirituality, the mind, physical fitness, extended family, life force in people and objects, unique identity of individuals and family, breath of life from forbears, and the open and healthy expression of emotion" (Ministry of Health NZ, 2017g). The third model of care attempts to include modern medicine into Māori health services. It is called the "Southern Cross Star Constellation" with four main stars representing cultural identity, physical environment, healthy lifestyles and participation in society (Ministry of Health NZ, 2017e). It also includes two "pointers"[8] to represent community leadership and autonomy to incorporate flexibility to address the changing demand for care.

A study on the implications of general care models on the workload of health professionals found that adopting models of care increased the workload on health professionals (Trindade and Pires de, 2013). Furthermore, a thematic review of the factors that influence a health care system into accepting models of care, found that most studies emphasised a patient's perspective and only a very few addressed a health care provider's perspective (Davy et al., 2015; Harris and Zwar, 2007; Walley et al., 2008). For patients, it found that the severity of their condition and the information available to them were major factors in their acceptance of new ways of care. On the

[8]These pointers refer to the two stars that "point" to the Southern Cross constellation.

other hand, health care providers were more open to accept new ways of care delivery provided they are confident in the ability of alternate provider roles to attend to the patient's needs (Harris et al., 2017).

There are models of care that encourage *self-directed care* where patients self-coordinate their care with other care providers based on their assessment of their own health. Although these models of care are proved to have better patient satisfaction, it heavily depends on the patient's ability to appropriately assess their own health needs.

The Guided Care Model (Hawkins and Novak, 2011) is one of the models of care that closely follows the CCM model. It involves a specially educated and trained nurse working collaboratively with multiple physicians to meet the needs of their chronic patients at an individual patient level. Although, Hawkins and Novak (2011) claims that this model of care was successful and effective, they did not mention the increased (or decreased) workload due to the changes in the practice. Likewise, most studies (Amir et al., 2015; Bodenheimer et al., 2005, 2002a; Dennis et al., 2008; Reeves et al., 2014) that refer to CCM, though they emphasise LTC management and focus on individual patient outcomes, do not mention the impact of adopting the model of care on the workload of health professionals.

Yet another model of care widely accepted is the UK's Quality and Outcomes Framework (QOF), introduced in 2004 (Downing et al., 2007). This framework primarily focused on achieving QOF scores, defined in terms of number of physician appointments, number of hospitalisation, etc., set as targets for ten specific conditions. The primary aim of this framework was to motivate GPs, through pay-for-service, to attend to longterm conditions. However, these QOF scores were not directly linked to adherence to guidelines, but rather to how health outcomes were recorded by the clinicians. Moreover, indicators considered those longterm conditions in isolation, and most recommendations on improving QOF suggested considering multimorbidity with a focus on population health (Downing et al., 2007; Forbes et al., 2017; Roland and Guthrie, 2016).

Although adopting one or more components of these models of care has improved the quality of care, with an emphasis on the requirement for training various roles as needed in the local context to attend to their population management needs, Davy et al. (2015) highlight the need of a system to predict the increase in workload due to adopting the right model of care for a medical practice. In this study, Davy et al. also discuss the importance of clinical information systems in adopting a model of care.

2.3.3 Simulation in health care

Simulation has been widely used in the health care domain. Simulation techniques used in health care include discrete event simulation (DES), system dynamics (SD) and agent-based simulation (ABS). In this section we discuss and compare these simulation techniques to our rule-based approach.

DES is applied in contexts where the environment (domain) can be represented as state variables, and simulation proceeds as these state variables change (Brailsford and Hilton, 2001; Law and Kelton, 1991). An advantage of using DES is improving efficiency of the simulation, as the internal clock of the simulation system can advance to a point where the next event is scheduled to occur (Zhang, 2018). Studies that apply DES include those of Eatock et al. (2015); Günal and Pidd (2010); Konrad et al. (2017) and Jun et al. (1999).

In our case, the medical practice has no data that reflects how well it adhered to guidelines in the past. Moreover, the patient information needed to apply guidelines is better suited to a knowledge-based representation than one using state variables. Therefore, DES is not an apt choice for our requirement of workload prediction from adherence to clinical guidelines. As discussed in Section 2.2.5, we unfold plan-of-care for each patient from the care plans. With an if-then format, care plans are rules in themselves. Therefore, a rule-based approach seemed appropriate in our context.

Systems dynamics (SD) is used to understand how a system changes over time, represented using "stock" and "flows" (Sweetser, 1999). In contrast to DES, SD models state changes as continuous processes. Therefore, SD is not suited to provide a detailed representation of a system where the state changes occur at discrete times (Brailsford and Hilton, 2001; Sweetser, 1999).

Agent-based simulation (ABS) is usually applied to understand the emergent behaviour of agents (Davidsson, 2001). This emergent behaviour may account for interactions between agents or interactions of agents with their environment. Either way, there is no requirement for agent-based simulation in our context, as we simulate the care pathways and not interactions between patients or health providers.

Shortliffe (1974) claims that the use of rule-based systems (RBS) in health care was started in 1973 through MYCIN. MYCIN is an artificially intelligent computer program that uses decision rules to help physicians make decisions regarding infectious disease treatment. It has a *consultation system*, an *explanation system* and a *rule-acquisition system*. The rule-acquisition system accepts and codes the rules for consultation from the physicians or experts. During an interaction with a physician,

its consultation system uses this rule-base along with the information provided by the physician, to recommend the clinical decision for the physician. The explanation system helps to answer the physician's questions regarding its decision (made by its consultation system). In MYCIN, each rule consists of a set of preconditions (called PREMISEs) which, if true, permit a conclusion to be made or an action to be taken, according to the ACTION part of the rule (Shortliffe, 1974; Shortliffe et al., 1975).

Fuzzy Cognitive Maps (FCMs) can be used to represent and reason with causality and uncertainty. FCMs are networks consisting of nodes and connections. The nodes represent clinical concept variables, such as observations and therapies and may take discrete or fuzzy values. The connections between the concepts and their strengths are drawn from clinical guidelines and are represented as ifthen rules that denote how one concept is used to infer another concept (i.e., a therapy option could be linked to its indications and contra-indications and fuzzy weights of the connections, in the range of 0 and 1, would mark the strength of evidence). The authors demonstrated their approach by modeling part of a urinary tract infection guideline through FCMs using semantic web tools. When patients data are input, the reasoning engine infers the values of activated nodes, ranking the different therapies. The modular representation of rules in FCMs could potentially facilitate knowledge reuse, sharing, and knowledge management.

While Papageorgiou (2011); Pawlak (1997) and Anooj (2012) use fuzzy (rule-based network) systems to make clinical decisions, Weiss et al. (1978) present a causal-associational network (CASNET) model. These rule-based network clinical decision support systems are flexible in terms of adherence to guidelines. A fuzzy rule-based clinical decision support systems has clinical concepts (e.g., observations or laboratory tests) connected to each other based on the the clinical guidelines represented as fuzzy inference rules (Anooj, 2012; Pawlak, 1997). When there are more than one connection between nodes, a measure such as the fuzzy weights or a probabilistic approach is taken to decide which path in the network to choose (Papageorgiou, 2011; Peleg, 2013). While in CASNET, observations of a patient are associated with a state of a disease, which is then linked to its classification. Because the observation can lead to various diseases, these states of disease and observations are represented as a network. However, this system is enriched with rules to associate states with observations and, to associate disease categories with states. Weiss et al. (1978) also illustrate how the CASNET model based system was used by ophthalmologists for long-term diagnosis and treatment of many types of glaucoma.

Pestotnik et al. (1996) embed clinical pathway decisions as rules in their clinical decision support system to implement decisions related to antibiotic prescriptions. They developed the rules thorough evaluations of published reports, use of national guidelines and local expert opinion, and exhaustive analyses of the LDS Hospital patient database. Although this system was deployed in a hospital setting, we follow a similar approach in the context of a PHC setting. We use the anonymised PHC patient data (of a medical practice) shared with us by our PCA; referred to national standard guidelines and local expert opinion to develop rules that unfold recall decisions from care pathways (care plans as some may refer to them).

A recent study by Azadmanjir et al. (2017) uses a rule-based system to develop a computer-aided coding system for the Iranian Classification of Health Interventions (IRCHI). Mabotuwana and Warren (2010) presents use of problem class that represent generic names for clinical encodings. Based on these studies, we map clinical encodings of classifications in our dataset to their generic names to support population-level reasoning.

Kahn et al. (1991) shows that reasoning methods (rules) can determine the temporal dependency between multiple clinical events. This study informs the strength of rules to embed temporal dependency required in our work (we require to anticipate "when" would a patient require a follow up appointment).

These studies show that, although the logic may be represented using different notations (e.g., fuzzy rules, network nodes and associations) rule-based systems are extensively used in clinical decision support systems. Moreover, a rule-base is scalable i.e., new rules can be added, modified or deleted; and, a rule-based system can use its rule-base to reason how and why it made the a decision (Archenaa and Anita, 2015; Davis et al., 1977; Krittanawong et al., 2018; Shortliffe et al., 1975). Although, initially rule-based systems were more uncertain on how to handle changes in facts or modified rules (Spiegelhalter and Knill-Jones, 1984), they have evolved through techniques allowing rule-based systems to react to the addition, modification or deletion of rules (Lhotska et al., 2001; Salatino et al., 2016).

We aim to predict workload from LTC management needs of patients at a PHC. We needed to predict the follow up appointments of LTC patients who follow (best practice) guidelines. We were informed, by the study of Archenaa and Anita (2015); Kim et al. (2018); Pestotnik et al. (1996) and Shiffman (1997), that implementing clinical guidelines as rules support the "if -then" structure embed in clinical decisions, and can be made adaptable to the changing context of the health care domain.

2.3.4 Workload management in health care

The Workforce Evidence-Based (WEB) planning model by Segal and Leach (2011) maps individual patient types to one of the following four levels: (1) promotion, prevention, and screening of the general or high-risk population; (2) type or stage of disease; (3) complications; and (4) threats to self-care capacity, depending on the severity of their health care needs. Given patient information, it suggests a unique clinical team to address each patient's health care needs. The authors argue that an organisation can, thus, estimate the total health force required. Following a similar approach as Seddon et al. (2001), we can analyse the financial and clinical risks in changing the practice's workforce. They also highlight the problems of quality of data and the difficulty in making decisions regarding how to combine care provider roles when patient attributes and health care needs vary across the population. They present involvement of clinicians and other care provider roles to define best practice for their population (diabetes in this case), which motivated us to follow the design science research approach (presented in the following chapter), which suggests involvement of various roles including GPs and chief medical officers in different phases of development and validation of our model.

The Working in Partnership Programme (WiPP), which included 13 initiatives, was developed by the NHS in the UK to support workload and capacity issues (Anonymous, 2006a). One of the initiatives was to have a workload analysis tool to improve workload management in general practice (Anonymous, 2006b). It was piloted in 60 GP surgeries, used to identify children with minor ailments or, to study the impact of including pharmacists to review LTC medications for their patients (Anonymous, 2006b). However, the details about how this toolkit was developed and deployed are not available for public access. Moreover, there were certain criteria in order to identify the pilot sites for WiPP programme (Prime, 2005), which suggests that the toolkit was not a generalised one.

Warwick and Bell (2007) propose a "Holon Framework" which has two modes ('soft' and 'hard') to address the soft W's (the What, Where and Who) as well as the hard elements (the Why, How and When) of health care planning. They showed that what-if scenarios can help to examine and enhance understanding of the effects of macro-management intervention in planning. Based on their approach, we chose to use what-if scenarios in our work to examine the impacts of alternatives in practices to meet the predicted workload.

The Care Management Event Tracking (CMET) feature of the Indian Health Ser-

vice's 2014 iCare project can track care events such as cervical screening, if a patient chooses to allow this. However, it can only track events related to breast, cervical, colon, liver, skeletal, prostate or STI events. Moreover, it is not clear what information will be presented (if any) if a patient chooses not to track the events, but a physician would like to track how many patients are postponing their care events. Furthermore, being a commercial product, it is not evident how the tool is developed. However, it is interesting to see from the tool's description that workload management issue is being addressed in different explicit and implicit ways.

Maher et al. (2009) suggest a framework that integrates chronic care services and has indicators that show the programme performance as well as how well the services are accessed in adherence to "standard" protocols of care management. They state that in order to deliver structured quality care, organisations need to establish a simple standard protocol for diagnosis, treatment and follow-ups. However, they do not address how these requirements can be met by the organisations. On the other hand, our tool can address such requirements; knowing the population and volume of sub-population visiting, the practice can plan and organise care actions accordingly.

The work of Massimi et al. (2017) and Busetto et al. (2017) emphasises the role of nurses in meeting the LTC management needs of their patients. Massimi et al. (2017) lists what interventions are possible by a nurse, and a review study by Busetto et al. (2017) found that most studies emphasised nurse involvement as part of their integrated care for LTC patients. We also examine the impact of shifting patients to nurses (task shifting) on the predicted workload. Moreover, there is a significant increase in the number of medications per day, visits to health professional and time spent for health-related activities, depending on the co-morbidity of the patients (du Vaure et al., 2016).

Table 2.6: Summary of the approach to develop our model and the studies that motivated this approach.

Our approach to develop our workload management model	**Studies that motivated the approach**
Workload prediction at a PHC centre is a known challenge	Abdel-Aal and Mangoud (1998); Green (2013); Green and Savin (2008); Harris and Zwar (2007); Murray and Berwick (2003); O'Leary et al. (2013); Potts et al. (2011); Reeves et al. (2014); Terry (2017); Trindade and Pires de (2013); Utley and Worthington (2012)

The volume of appointments or care activities is considered as the workload of a PHC	Amir et al. (2015); Anonymous (2006b); Burt et al. (2014); Fox et al. (2006, 2009); Hall (2012); Heroman et al. (2012); Jordan et al. (2003); Mathers et al. (2011); Ministry of Health NZ (2017k)
Recalls of a patient can be predicted by analysing care pathways, care plans or treatment plans of patients	Best Practice Advocacy Centre New Zealand (1997); Brown et al. (2018); Davy et al. (2015); Fox et al. (2006); Jackson et al. (2013); Kahn et al. (1991); Mathers et al. (2011); Stokes et al. (2017)
Multi-disciplinary team can address workload better	Åhgren (2003); Davy et al. (2015); Grant and Greene (2012); Jackson et al. (2013); Pines et al. (2015); Reid et al. (2009); Skipper (2010); Stokes et al. (2017); Struijs et al. (2015); Taylor (2015); WHO (2016b)
Ability to foresee the upcoming workload can help analyse and improve efficiency and effectiveness of care delivered	Anonymous (2006a); Bodenheimer et al. (2009); Davy et al. (2015); Harris et al. (2017); Seddon et al. (2001); Segal and Leach (2011); Wagner (1995)
A rule-based approach would be most appropriate to implement care pathways	Archenaa and Anita (2015); Azadmanjir et al. (2017); Kim et al. (2018); Mabotuwana and Warren (2010); Pestotnik et al. (1996); Shiffman (1997)

In summary, delivering effective and efficient chronic care at a PHC level has been challenging for decades. Bodenheimer et al. (2002a,b); Wagner (1998) and Ham (2010) discuss the changes that are required to be a high performing chronic care system. We outlined a few case studies that incorporated one or more CCM components in their care delivery. We discussed the CCM components, specifically those related to the features of our model. We also saw that although the emphasis of CCM components varies across studies, most of them had more than one CCM component addressed in their work. Then, we discussed those works which either discuss the workload at a PHC, how services are planned and delivered, various simulation techniques, and the workload management strategies that relate to our work in some way.

In brief, we use a rule-based approach to implement care pathways and simulate

patient visits depicted as the workload that needs to be managed within the capacity of the practice. Table 2.6 highlights the related work that motivated our approach. In our discussion in Section 9.1, Table 9.1 gives an extensive comparison of our work to related work discussed here.

Chapter 3

Research Model and Methodology

Our research is partially motivated by the requirement from our PCA, whose role involved planning care and resources at an organisational level. According to the PCA, he needed some "way" to measure the volume of work that needed to be planned. He, being a part of an organisation that promotes best practice in PHC in New Zealand, also needed a tool that would help the practices understand the impacts adopting best practice of care delivery. This requirement for gathering data and building awareness of the problem in the PHC domain initiated our design science research (DSR) to develop our Adaptable Best Practice based Workload Prediction Model (ABP-WPM).

According to March and Smith (1995) in a DSR approach much emphasis is on that the research outputs serve human purposes. In that study, March and Smith present a research framework for information technology, that separates research activities from research outputs. They claim that in a design science research (DSR), research activities (build and evaluate) create the research outputs (constructs, models, methods and implementations) using a 'technology oriented' approach.

As aforementioned, according to March and Smith (1995), constructs, models, methods, and implementations are the four types of research outputs. Constructs are the concepts that are used to define problems, the phenomenon or specify solutions in a domain. These constructs (concepts) can be combined to form a model. Such a model would describe the tasks, activities or processes. For example, variables in a mathematical equation form the constructs and the equation can be a model and the operators used in the equation is the process. In other words, a model describes the relationships between constructs. They also say that natural science researchers use *theory* as a synonym for *model*.

March and Smith also consider that design scientists may develop methods to

achieve or describe goal-oriented activities. Thus, methods are based on constructs (or the concepts) and a model (representation of the solutions). Finally, developing specific products or instantiations of a process can also be considered as an outcome of DSR. They emphasise that "rather than posing theories, design scientists strive to create models, methods and implementations that are innovative and valuable" (p.254).

The research activities set out by March and Smith also focus on building an artefact and demonstrating that such an artefact can be built. Evaluation refers to assessing the performance of the developed artefact against some (pre-identified) criteria. According to them, "theories explicate the characteristics of the artefact and its interaction with the environment that result in the observed performance" (p.259). In that way, they consider that theory building and justifying that theory are also research activities. The following section describes our research model in these aspects.

3.1 Our model

3.1.1 The three-layer LTC PHC construct

Our aim is to help medical practices foresee the workload from their LTC patients. Therefore, in our domain we focus on LTCs, LTC patients and LTC population (refer to Figure 3.1). A care pathway sets guidelines to manage an LTC, and a care plan is an instantiation of the care pathway for that LTC in a patient (see Level 1: Instantiations in Figure 3.1). Multiple LTCs in a patient means that multiple care plans applied are to a patient. We then merge these care plans applicable to a patient as an individual plan-of-care (this is depicted as Level 2: Care plan Aggregation), which then is used to predict the workload from the LTC population (depicted as Level 3: Abstraction and Aggregation). In terms of our DSR outputs, we have a three layer LTC PHC construct to predict workload due to LTC patients at a PHC level.

3.1.2 The encounter-based unfolding plan-of-care process

Figure 3.2 presents our research model. During a patient visit, based on the patient information (in the figure, the dashed line from PHC patient data shows this dependency), our model first creates disease specific individual care plans. Our rules then merge these disease specific individual care plans to create individual plans-of-care (this considers multi-morbidity in a patient), and the next intervention required for the patient is planned. These individual plans-of-care are then aggregated to create

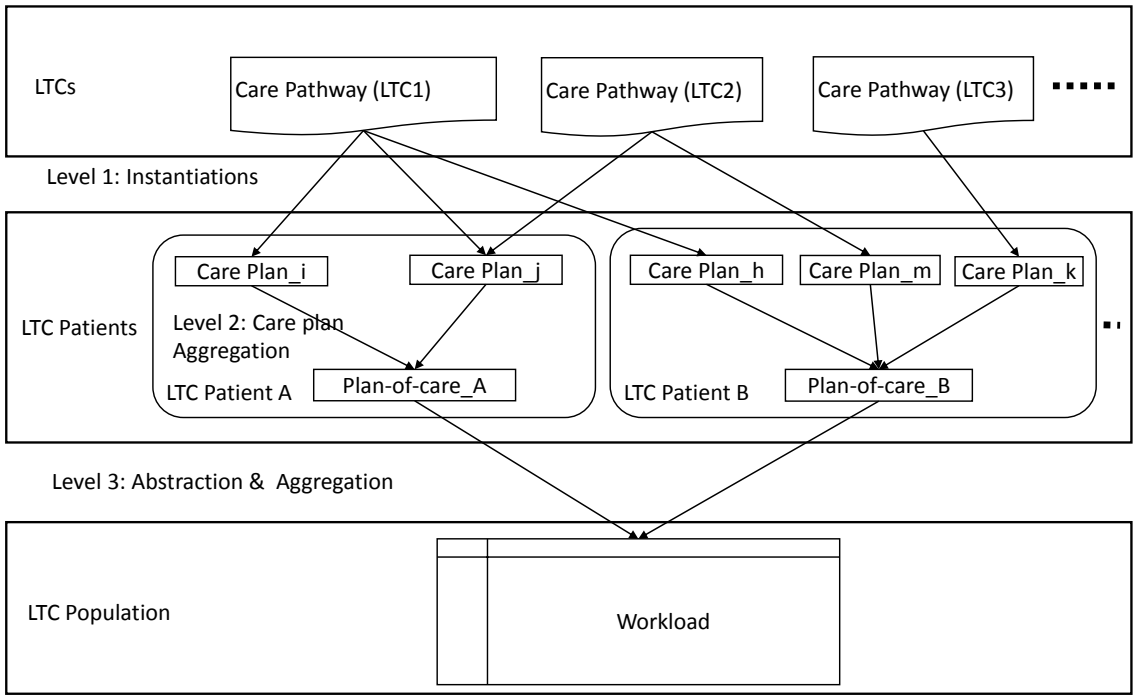


Figure 3.1: Theoretical constructs used in our model.

the population-level plan-of-care. Thus, the care for an LTC patient, as required by the best practice guidelines, "unfolds" visit by visit of an LTC patient. Because this phenomenon occurs at patient visits, and unfolds care during each visit, we call this process an "encounter-based unfolding plan-of-care" process. Our PCA advised that this is the way in which GPs experience plan-of-care.

3.1.3 The Adaptable Best Practice based Workload Prediction Model (ABP-WPM)

Our workload prediction model is based on the above theoretical constructs, and follows the encounter-based unfolding plan-of-care process to predict the population level workload at the primary health centre. Within our workload prediction model, we encoded these care pathways (the best practice guidelines shared by our advisor) as rules. This enhances our workload prediction model as an adaptable best practice based workload prediction model (ABP-WPM).

We evaluated whether this aggregated plan-of-care is of use by applying it (through simulation) to predict the workload from the LTC patients at a PHC centre. We found

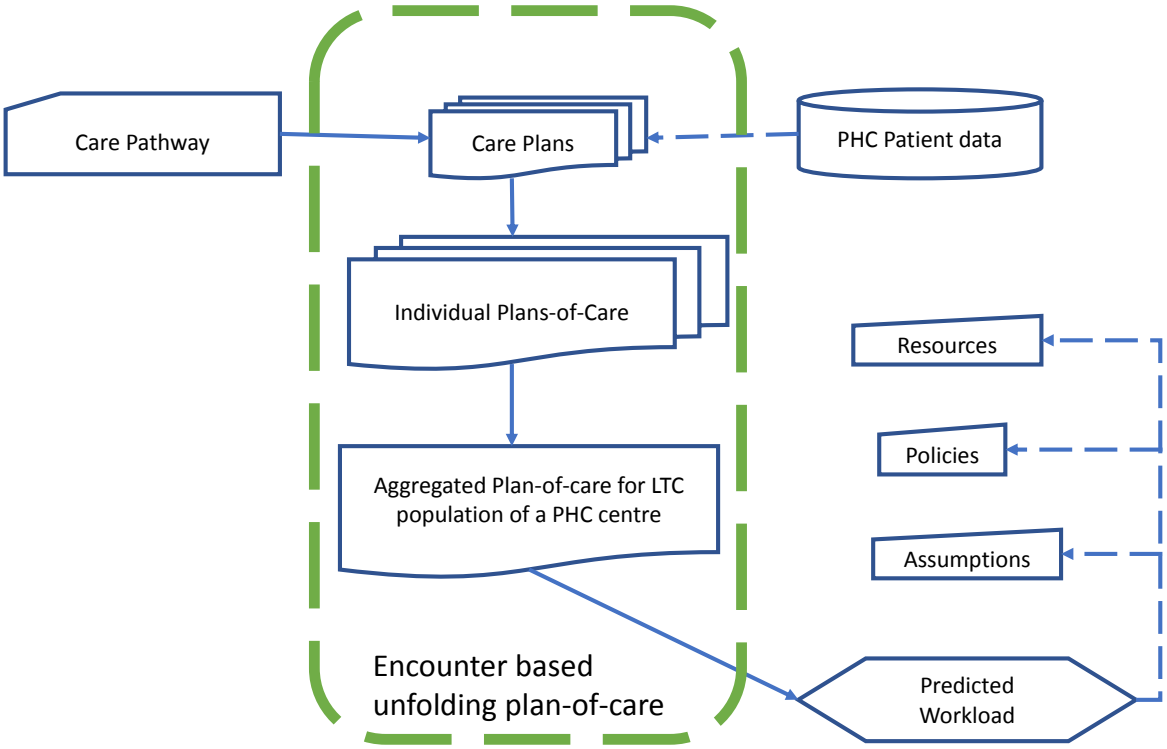


Figure 3.2: The encounter based unfolding plan-of-care process.

that it can help the primary health care centre to foresee their upcoming workload and plan accordingly. Agreeing with March and Smith (1995), this instantiation of our model to predict the workload, demonstrates the utility and usability of our three layer LTC PHC construct, encounter-based unfolding plan-of-care process as well as our model.

Thus, our domain is a primary health care centre. We use a three layer LTC PHC construct with care pathways, care plans and individual plans-of-care for LTCs, LTC patients and LTC population, respectively, in a primary health care context. We built our workload prediction model (i.e., the artefact) which follows an encounter based unfolding plan-of-care to create an aggregated plan-of-care for LTC patients. We evaluate this plan-of-care by applying it to predict the upcoming workload from the LTC patients at a PHC centre. This workload then informs the resources, policies and assumptions and update the rule-base accordingly.

3.2 Our DSRM approach

The work presented in this thesis follows the *design science research* methodology (DSRM). DSRM comprises iterative steps involving identifying (and motivating) a problem, defining the desired properties of a solution, designing and developing an artefact (software in this case), and evaluating the developed prototype through various iterations of system development (Baskerville et al., 2009; Hevner, 2007; Vaishnavi and Kuechler, 2007). An important step, which distinguishes DSRM from solution development, is identifying the (general) lessons learned, communicating them, and using the feedback from a broader community for the subsequent iterations of development of the artefact. The process is informed by relevant literature as stated by Peffers et al. (2007): "development of the artefact should be a search process that draws from existing theories and knowledge to come up with a solution to a defined problem" (p. 48). As aforementioned, DSRM requires one to communicate and collect feedback on the prototype developed from a broader community; for this work, the broader community included our Primary Care Advisor of this work, members including GPs, health operation officers, and chief medical advisers for the Waikato PHO and the various care providers at Mosgiel PHC centre.

3.2.1 The holistic view of our research

As presented in our research questions in Section 1.2, the main aim of our work is to find what model(s) of care can help manage the workload at a PHC centre. We initially started focusing on the care activities at a PHC centre, which later required us to identify specific patient care needs. As mentioned in Section 1.3, we used a de-identified dataset of a PHC centre.

Our major challenges were due to

1. specific clinical encodings in health data and missing information, such as whether a patient is a longterm patient or not, added more complexity. For example, C109 and C108 are Read codes that represent diabetes in a patient, and the clinical guidelines do not reflect explicitly which Read code corresponds to the path in the clinical guideline.

2. inconsistent usage of terminologies in health literature; although in this work, we have a clear distinction between the terms used for planning care, as discussed in Section 2.1.1, the health care domain uses these terms interchangeably.

3. not having a complete understanding of how care planning for LTC patients is performed in a real-world scenario. The clinical guidelines do not explicitly represent which paths to follow for different Read codes. Therefore, we primarily relied on the feedback from our PCA to map the health care knowledge and decision-making process to an IT domain.

These challenges initiated our cyclic approach to develop our ABP-WPM. This research follows an intertwined three-phase process. Figure 3.3 shows a timeline for these three phases, namely the information interpretation (II), the rule-base development (RBD) and the phase to improve simulation capability (SC).

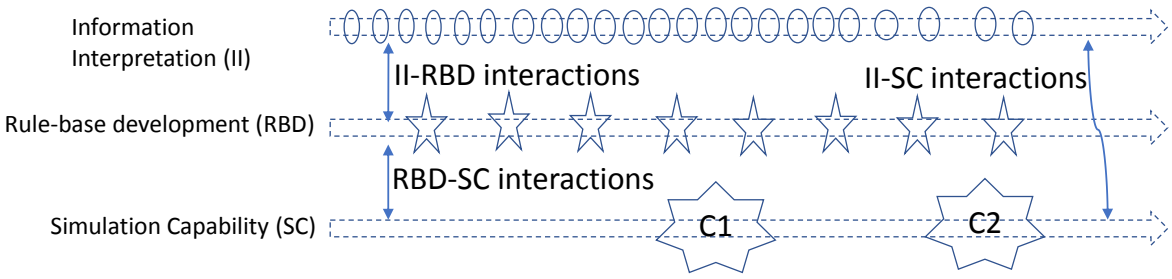


Figure 3.3: The holistic view of our research process.

Consider the SC time line. The stars, named C1, C2 and so on, are the iterative cycles (presented in Section 3.2.2) for improving simulation capability (SC). These require interactions with the rule-base development (RBD) phase. The RBD phase developed the simulation rules, which in turn, were evaluated in the context of each cycle in the SC time line. These interactions (although the figure shows only one line, there are more and were frequent interactions) are shown using *RBD-SC interactions* line.

Furthermore, the rule-base was also impacted by the information in the patient data. And, due to the aforementioned challenges, this required us to undergo rigorous iterative feedback cycles within the II phase. The feedback of some of these II-II cycles were fed into the RBD phase. Thus, apart from the more frequent II-II cycles (represented by the circles on the II time line), there were also interactions with the RBD phase (represented by the *II-RBD interactions* line). Because the information required depended on the SC context, at times these II-II interactions directly impacted our SC phase, which is depicted the *II-SC interactions* line. Chapter 8 presents these interactions related to the II phase.

For example, we needed to know how many patients are recalled every three months for diabetes. In order to have this capability (SC), we needed to understand how dia-

betes patients can be identified from the data. There were more frequent II-II interactions in order to understand the required information such as what clinical encodings were used in the data and how these clinical encodings can be mapped to identify the diabetic patients. This feedback help us understand what care plans mean in the context of deciding the recall period (II-RBD interactions) and how rules can be formulated (RBD-RBD interactions). In order to decide recalls in case of multiple conditions in diabetic patients (RBD-SC interactions), the information of which were fed back to the II phase. Thus, our rule base was developed in an iterative feedback driven manner as described in the following section. Please do bear in mind that although we explain these iterative cycles separately, in reality these cycles are intertwined. For now, we will focus on the iterative cycles of development of the workload prediction phase of our ABP-WPM.

3.2.2 The iterative-cycle approach

The DSRM approach as presented by Vaishnavi and Kuechler (2007) comprises six iterative steps. Figure 3.4 shows our DSRM process adapted to follow the DSR process by Vaishnavi and Kuechler (2007).

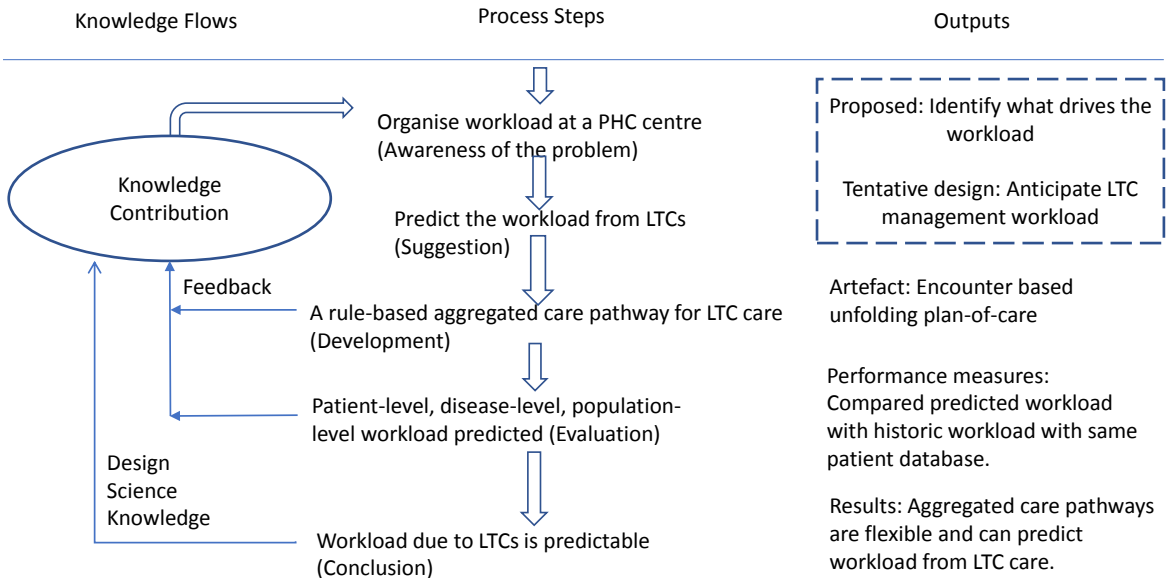


Figure 3.4: Our DSRM process (adapted from DSR process presented by Vaishnavi and Kuechler (2007) (p. 15).

Our work is to help a PHC organise its workload (i.e., the awareness of the problem). Therefore, our initial proposal was to identify what contributes to this workload. We,

then, identified that growing demand for care due to LTCs pose a major challenge to manage their workload. Hence, our tentative design involved anticipating the workload due to LTC management needs. Therefore, our first step is to predict LTC load on primary health care. Having the problem identified, the next step was to support LTC workload prediction at a population level. This involved identifying and stratifying the patient population based on LTCs present in them. In order to do so, we developed a rule-based system that simulates and aggregates the LTC patients' individual plans-of-care to a plan-of-care i.e., the encounter based unfolding plan-of-care for the LTC population at the PHC centre. This aggregated care pathway, for LTC care needs, when applied to the patient data can anticipate the care needs at a patient-level, disease-level and population-level (as discussed below). This demonstrates that the workload at a PHC centre is predictable. Specifically the knowledge contributions are

1. the individual plans-of-care can be aggregated into a single care pathway for LTC population using a rule-based approach,

2. the workload at a patient-level, disease-level and population level is predictable.

The final steps of iteration evaluated the model developed by comparing the predicted workload to the actual historic workload from the same LTC patients.

The goodness of fit of a predictive model describes how well it fits a set of observations. However, in this work, we do not have such a set of observations with which to compare the goodness of fit, as the historic workload does not comply to best practice guidelines and we aim to simulate the workload arising from adherence to these guidelines. Similarly, it is not useful to measure the difference (error) between these simulated and historic workloads, which rules out applying standard statistical measures such as RMSE for comparison of simulated workload. We expect the two workloads to differ, and a numeric measure of the difference would not help to validate our model, as there is no level of difference that is known a priori.

Dahabreh et al. (2017) defines "face validity" to refer to "whether the model is deemed a satisfactory representation of the salient aspects of reality and whether the model results appear to be plausible". Some recent studies such as those of Connell et al. (2018); Jorgensen et al. (2018) and Luke et al. (2018) use face validity as a measure to evaluate models, where usefulness is based on the understanding of a group of experts in the domain. Similarly, Gacenga et al. (2012) and Leukel et al. (2014) uses "expert evaluation" instead of the term face validity to refer to the DSR phases of validation and evaluation of their frameworks.

In this work we also chose to evaluate our model using expert evaluation. At various stages of development of our model, the results were communicated and feedback was collected from our PCA, the Planning and Funding Manager of Waitemata DHB, the members of Waikato PHO, the health professionals at Mosgiel Health Centre and the executive members of South Island PHO. Based on the feedback, the model was extended to include more LTC care pathways, and also corrected (tuned) to adapt the best practice guidelines for managing these LTCs in a patient. The model was further evaluated for the ease of predicting workload under various what-if scenarios (as presented in Chapter 6).

Figure 3.5 shows the iterations at a macro-level (the RBD - SC interactions in Figure 3.3, although there are inputs from the II phase as presented in Chapter 8). Please be informed that in practice there were micro-level iterations within steps A, B and C of each cycle (as depicted in Figure 3.3).

Chronic Kidney Disease (CKD) is one among the various LTCs that is believed to be stable in a patient for a year (National Kidney Foundation, 2015). Hence, we initially looked at the CKD care pathway (see Appendix A.4). In order to address cases of multi-morbidity, we extended the model to include patients with diabetes. In the following cycle, we included the wider population of LTC patients by including hypertensive patients. The final model addressed the complete LTC patient population at the practice. This predicted workload is then used to analyse various what-if scenarios. The same cohort of patients is used to implement our Bayesian inference. The rest of this section will detail these macro-level cycles and how the feedback from each cycle drove/set requirements for the cycle that follows it.

Cycle 1: Develop the patient recall rule engine for the CKD patients only.

Chronic Kidney Disease (CKD) care plans were decided based on the CKD stage in a patient. The challenge was to identify the CKD stage in a patient, given that it is not identified by Read codes but by their eGFR and albumin-creatinine ratio (ACR) laboratory values (presented in Section 4.5.2). I take this opportunity to remind the reader that this study does not consider changes in the plan-of-care for a patient during the simulation period. Therefore once the CKD stage in a patient is identified, the recall period for the patient due to CKD is fixed for the simulation period.

Having the care pathway for CKD implemented, we extended the rule-base to simulate patient visits and schedule corresponding recalls[1] (See Section 5.2 for the complete

[1]In this work, subsequent appointments for LTC management needs are referred to as recalls.

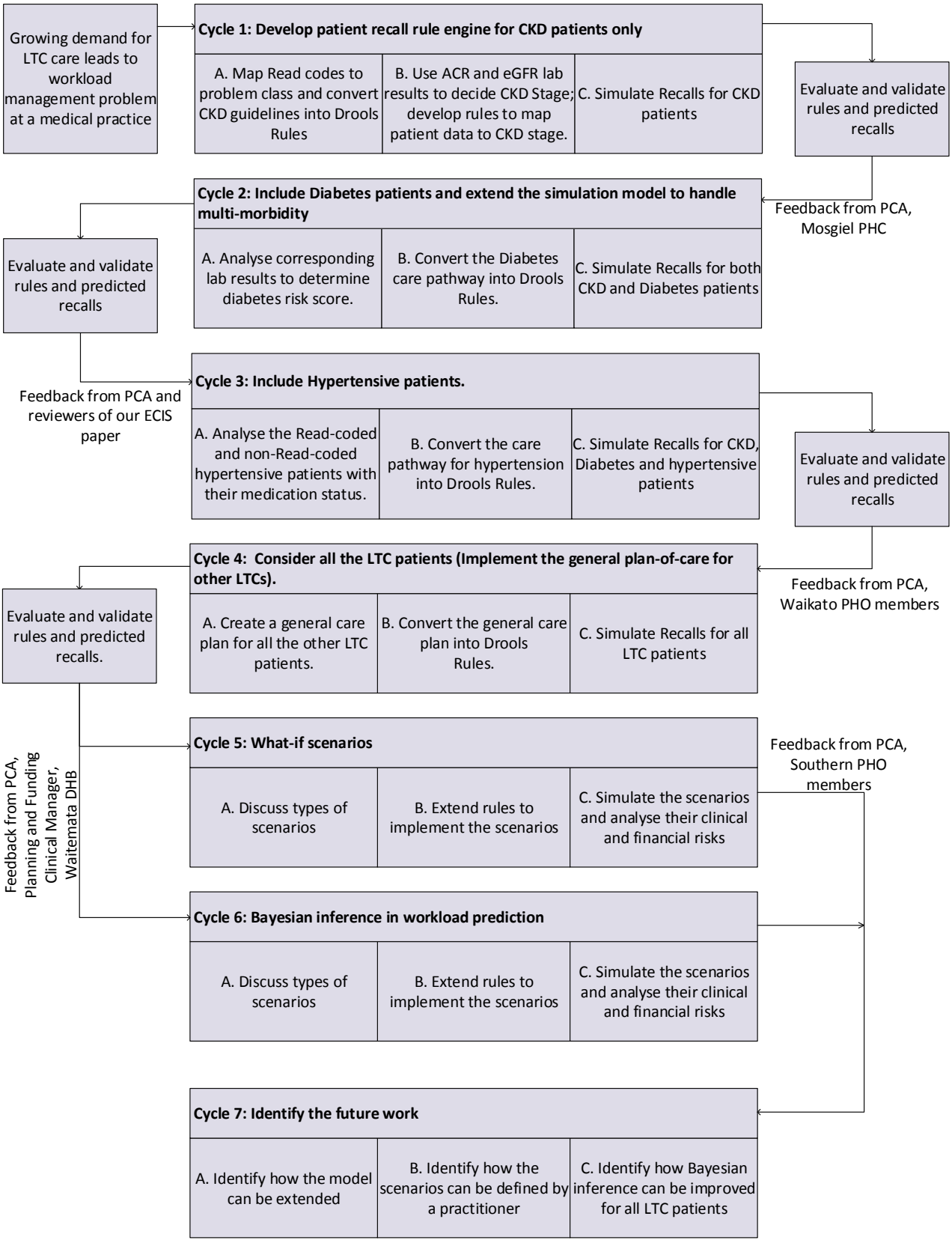


Figure 3.5: Iterative cycles followed to develop our rule-based workload simulation model.

Table 3.1: Multi-morbidity matrix.

Care plan code	asthma	Atrial	**CKD**	COPD	**Diabetes**	heartdisease	heartfailure	**Hypertension**	PAD	psychosis	stroke
asthma	1406	52	177	45	92	83	20	174	5	8	18
Atrial	52	393	253	43	68	148	73	114	20	2	42
CKD	177	253	**1498**	130	342	406	131	0	62	8	102
COPD	45	43	130	288	50	81	20	78	19	1	18
Diabetes	92	68	342	50	**640**	159	45	0	32	7	42
heartdisease	83	148	406	81	159	764	93	281	43	4	60
heartfailure	20	73	131	20	45	93	162	22	15	2	15
Hypertension	174	114	0	78	0	281	22	**1399**	24	9	67
PAD	5	20	62	19	32	43	15	24	97	0	15
psychosis	8	2	8	1	7	4	2	9	0	46	1
stroke	18	42	102	18	42	60	15	67	15	1	190

discussion of the rule-base for our simulation). Since only one condition was considered to simulate the workload, it was unlikely to be accurate. However, the findings and the feedback on the findings were used to improve the workload simulation as discussed in Chapter 4, Section 4.6.2. Knowing that the workload from the CKD patients was predictable, the rule-base was, then, extended to address the multi-morbidity in patients. In order to get predictions right, we need to consider all conditions (in case patient is recalled more frequently due to another condition). And, for the cohort of patients with diabetes, the most commonly recurring co-mortality is CKD (see Table 3.1, out of 640 diabetic patients, 324 has CKD). Hence, we chose to implement the best practice guidelines for diabetes.

Cycle 2: Extend the simulation model to include diabetes patients.

The care pathway for diabetes shared by the PCA was converted into Drools rules (presented in Section 4.5.3) and added to the rule-base for the workload simulation. Although the predicted workload data looks appealing, there were peaks and dips in the predicted workload which were not evident in the historic visits of these patients, as shown in Figure 3.6 which depicts the recalls for 3-monthly recalled patients.

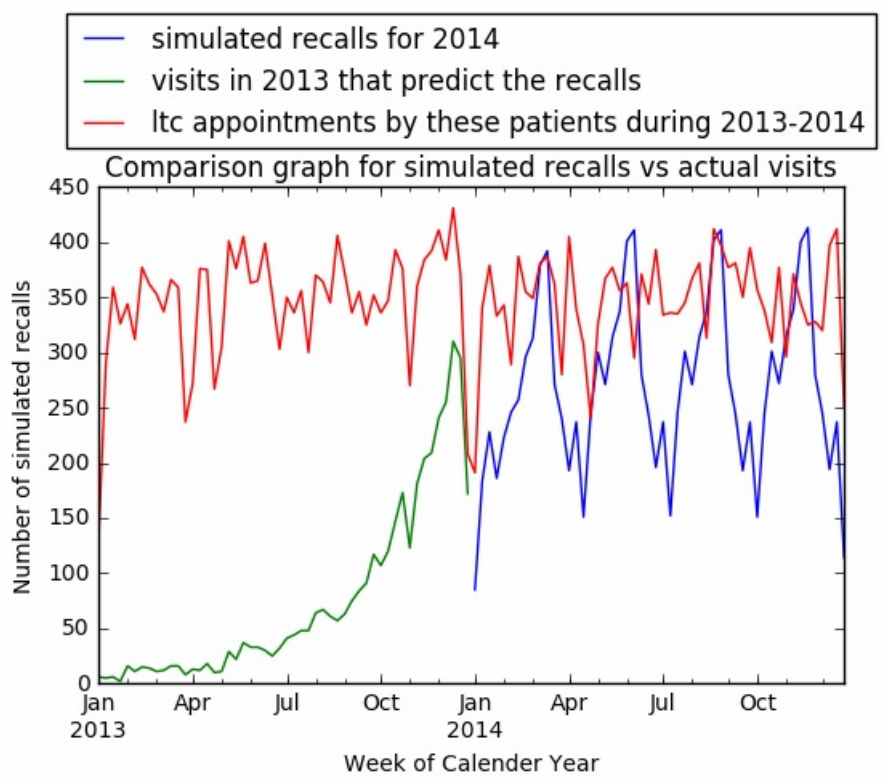


Figure 3.6: Initial results of simulation for 3-monthly recalled patients.

As evident from the figure, there is periodicity in our simulated recalls. This could be due to most recent visits in 2013 accumulating towards the end of the year. Another reason could be the presence of other LTCs meaning that the patients would be recalled more frequently than required by their CKD and diabetes care plans. Based on the multi-morbidity table, prevalence of hypertension in patients suggests that the workload from co-morbid patients could be more accurate if the care plans for hypertension was also implemented. Initially, our plan was to apply hypertension care plan to all the hypertensive Read-coded patients. However, the guidelines to manage hypertension recommended care actions for those who are not Read-coded with diabetes or CKD (the first line of care pathway for hypertensive patients given in Appendix A.6 recommends this. Therefore, in Table 3.1 it shows zero for the row CKD and column for hypertensive patients, and row for diabetic and column for hypertensive patients). The reviewers of our ECIS paper, (Devananda et al., 2017), were also keen to see the implementation details of the care pathway for hypertension. Hence, the next cycle included the care pathway for hypertension to simulate the workload due to LTCs.

Cycle 3: Extend the simulation model to include hypertensive patients

Most often, hypertension is not considered as an LTC, but a consequence of LTCs in a patient (Sarafidis et al., 2017). There is an on-going debate among general practitioners on acceptance of a care pathway for hypertension, especially in a PHC setting (Borgmeyer, 2013). The complexity of applying the care pathway for hypertension in the context of workload prediction is that a few GPs do not record hypertension for the patients, as they consider it as an exacerbation of one's blood pressure due to some factors. Therefore, there are two categories of hypertensive patients in our data: the Read-coded and the non-Read-coded hypertensive patients.

Another aspect that complicates delivering care was planning care for patients with extreme blood pressure (BP) readings. For such patients, intensification of care involved frequent visits (e.g., weekly or two-weekly recalls) and frequent medications to control the extreme blood pressure readings of a patient (Sarafidis et al., 2017). However, due the complexity of identification and application of intensified care for such patients, this study does not consider intensification as a part of the normal PHC workload. We consider intensification of care for hypertensive patients as a case for special care similar to an emergency care or the after-hours care.

Multiple BP readings for the same patient at different times of the day added more complexity to decide which care plan of the hypertensive care pathway should

be applied to a patient. In such cases, guided by the Primary Care Advisor, this work considers the most recent BP reading for the patient.

As aforementioned, there were Read-coded and non-Read-coded hypertensive patients. However, irrespective of whether the patient is Read-coded or not, the GPs prescribe medications for controlling the hypertension in their patients (Sarafidis et al., 2017). Hence, in this work, a factor that determined inclusion of a patient in the hypertensive patient cohort was their medication status. The subsequent feedback cycle, therefore, focussed on defining the medications, such as diuretics and non-diuretics medications, and how long they are on these medications, to help identify the severity of hypertension in them. This, in turn, decides the care plan applicable to a patient. The resulting care pathway for hypertensive patients is given in Appendix A.6.

The resulting predicted workload from the hypertensive patients (refer to Figure 3.7) seemed reasonable (as for hypertensive patients medications to control extreme BP values play an import role in their more frequent visits, which we do not address due to its complexity at an individual patient level) to our PCA, although the periodicity in simulated recalls (the red line in Figure 3.7) was not completely resolved.

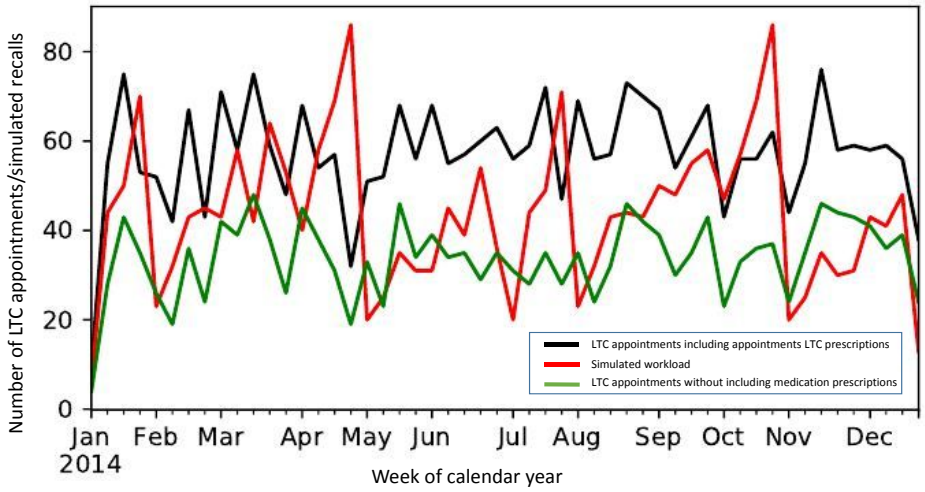


Figure 3.7: Comparison of simulated workload (the red line) with historic LTC visits (including visits for medication prescription (the black line) and without medication prescription (the green line)) by the hypertensive patients.

The predicted workload, these challenges and assumptions were discussed with members at the Waikato PHO which included various roles such as GPs, health operation officers, and chief medical advisers for the Waikato PHO. They also gave an insight into realistic assumptions as discussed in the following cycle. Having addressed the multi-morbidity, complexities of LTC patients and delivering care using care path-

ways, we were then interested to include all the LTC patients for our workload simulation. Extending our model to include all the LTC patients also, thus, tested the scalability of our rule-base.

Cycle 4: Extend the model to include all the LTC patients

There was no explicit information regarding the LTC status of a patient. Hence, we had a workaround to identify LTC patients and their corresponding appointments (see Section 4.4.1) to drive our workload simulation. As the scope of this work is limited to producing a proof-of-concept workload model, it is not necessary to implement care pathways of all LTCs. However, in order to include most LTC patients, agreeing with O'Halloran et al. (2004), we consider a six-monthly recall for general care plan reasonable for patients with other LTCs, since the vast majority[2]. of cohort patients have CKD, diabetes or are hypertensive. This base care plan aligns with the New Zealand's Ministry of Health Strategy to prevent, identify or early diagnosis of other LTCs (Ministry of Health NZ, 2016b). The generated workload was then compared to the historic LTC visits by these patients. According to our Primary Care Advisor on this work, Figure 3.8 comparing the generated workload with historic visits by these patients is reasonable.

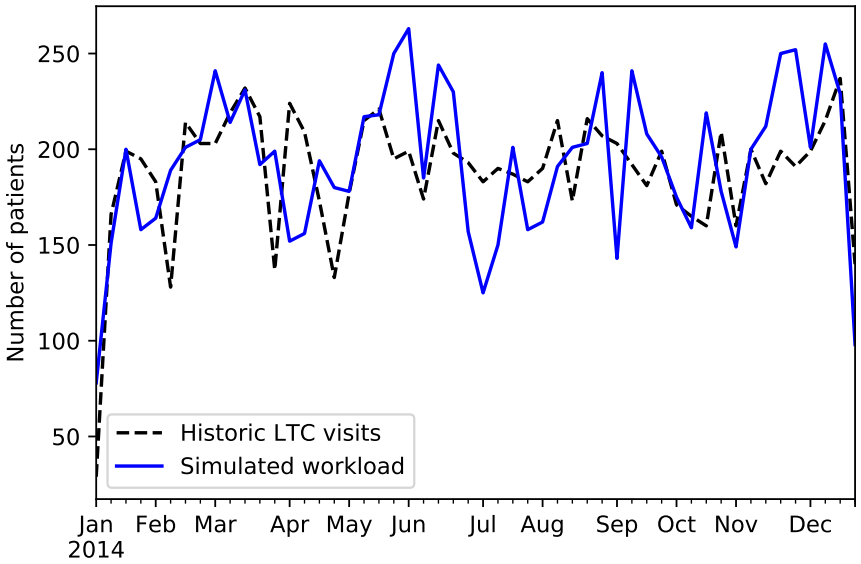


Figure 3.8: Comparison of simulated workload with historic LTC visits by the same patients.

[2]Out of 9 576 LTC patients, 7 047 patients were classified with at least one of the LTCs: CKD, diabetes or hypertension

After Cycle 4, the suggestions from our PCA, the Planning and Funding manager of Waitemata DHB, and the healthcare decision makers of Waikato PHO were the following.

- We should try to learn, in some way, the patients' visit patterns;
- We should consider shifting workload to nurses. It was also recommended that we could use severity of LTCs as a parameter to decide which patients could be attended by nurses;
- They also observed that usually when a policy changes, the health professionals are well aware of the clinical risks, so recommended that it would be valuable to address the financial risks even if it is a basic model.

Therefore, we proceeded with Cycle 5 and Cycle 6 in parallel, which both depend on the outputs of Cycle 4. Cycle 7 considers feedback from PCA and Southern PHO members and gives an insight to the usability of what-if scenarios and future directions. Thus, I added Cycle 7 as a phase of DSR that closes the loop, discussing the scope for future work.

Cycle 5: What-if scenarios

First, I explore the alternatives of practices that can be deployed, knowing the upcoming workload, at this PHC centre. In an informal meeting with the Planning and Funding Clinical Manager for Primary Care of Waitemata DHB, he suggested to have an analysis from a financial perspective would be good. He reflected, with his experience with PHOs and DHBs, that usually when a health scheme is recommended, what the clinicians need is to know how it impacts them financially (because the clinicians think the clinical risks can be handled by them). From a DSRM perspective, this step had more micro-level cycles of iterations to decide which scenarios to choose and how the financial aspects should be considered. In this exploratory study, I cover both the clinical and financial impacts of a few such alternative practices as discussed in Chapter 6.

Cycle 6: Bayesian inference in workload prediction

Second, we agree that our rule-base model (presented till now) assumes that patients turn up promptly (more rigidly) for their recalls when their medications run out. However, the appointments data of these patients reveal that it is not the case. Therefore,

we apply the Bayesian inference to learn about the patient visits and, further use that learning in our workload simulation (presented in Chapter 7). This step was guided by statistical advice from the experts from the Department of Statistics and Mathematics of our university.

Cycle 7: Identify the future work

We presented our work to the Southern PHO members and our PCA. There was feedback on how we can extend the model for future work. Based on Cycles 4 and 5, we also need to identify how the rule-base can be extended giving the practitioner the autonomy to define the what-if scenario they need to analyse in their context. Based on the output from Cycle 6, we ourselves identified that a future work would address incorporating Bayesian inference for all (LTC) patients.

3.3 Our contributions

Table 3.2: Comparison of our ABP-WPM to other uses of care pathways.

Our DSR outputs	**Characteristics**
Three layer LTC PHC constructs	Layered approach to decide plan-of-care for a population of LTC patients. Specific focus on LTC context.
Encounter-based unfolding plan-of-care process	Unfolds care during each visit of an LTC patient. In our case, we follow a dynamic approach of merging care plans during a patient visit.
Adaptable best practice based workload prediction model (ABP-WPM)	Three layered LTC PHC constructs. Follows the encounter based unfolding plan-of-care process. Can predict the upcoming workload from LTC management needs of patients at a PHC level.

Following DSRM, we make three main contributions. Firstly, the three-layer LTC PHC construct addresses the requirement of a terminologies and the relationship between them to plan LTC care at a population level. Secondly, we contribute to the

process of transferring care requirements from disease specific care pathways to a population of LTC patients at a PHC through our encounter-based unfolding plan-of-care process. Finally, our adaptable best practice based workload prediction model (ABP-WPM) that given patient information, can predict the upcoming workload, for a specified period, at a PHC centre. Table 3.2 summarises characteristics of our DSR outputs.

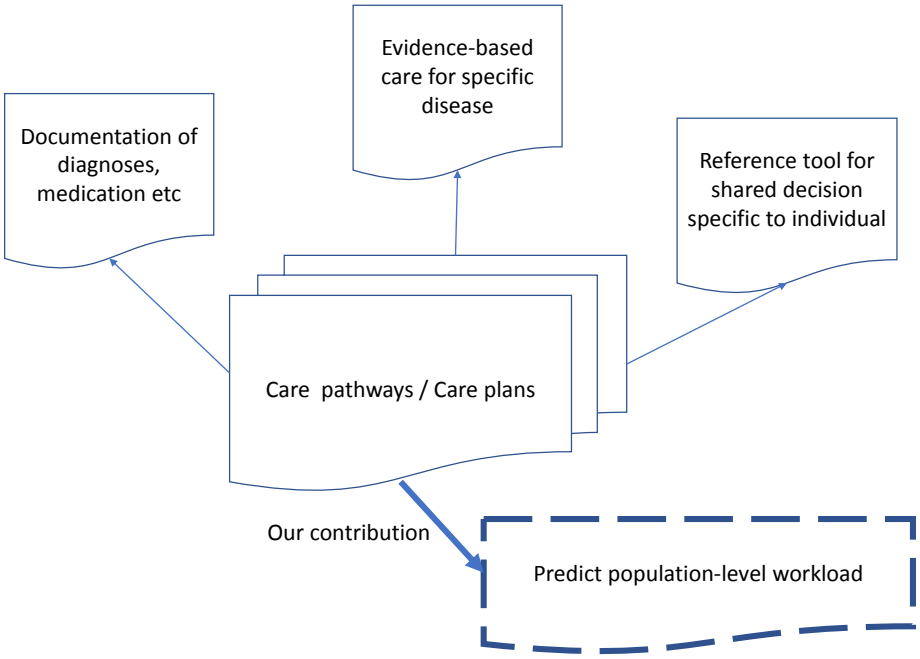


Figure 3.9: Our contribution to using care pathways.

In a broad sense, we add to knowledge on how care pathways can be used to predict upcoming LTC workload. As presented in Chapter 2, care pathways[3] are used in various contexts, we contribute to this list through our ABP-WPM as shown in Figure 3.9. As discussed in Section 2.1.1 care pathways, presented as flowcharts (NICE, 1999), are disease specific and guide to manage that specific disease in any patient (Best Practice Advocacy Centre New Zealand, 2012; NICE, 1999). Practices adhering to these care pathways reduces variance in providing care for a specific disease (Burt et al., 2014). However, in isolation, they may cause fragmented care in the case of multi-morbid patients. Some studies such as those by Ash et al. (2000); Busetto et al. (2017) and Fox et al. (2009) refer to care plans (usually in the form of electronic health records

[3]In literature, care pathways, care plans and treatment plans (or plans-of-care) are used interchangeably. Here I use this term to refer to care delivered to patients.

(EHRs)) for a shared decision for an individual. However, they do not use care plans (in any form) apart to make shared decisions for an individual. A few studies (e.g., (Babiker et al., 2014; Bauman et al., 2003)) use care plans to record a patient's health condition (similar to in EHRs). Sometime these documentations could be paper-based records. As the data recorded is specific to who provides care, chances are that the data recorded may be incomplete. In comparison to the above uses, our ABP-WPM is best practice based (uses care pathways and is adaptable based on patient needs), flexible (our rule base is extendible as required) and can be used to predict the upcoming workload from a group of patients. Currently, the scope of our work is a PHC context.

Having discussed our research model and methodology, in the next chapter, I present the rules in our rule engine that enacts on our three layer LTC PHC construct and follows the encounter based unfolding plan-of-care process, and the initialisation of our ABP-WPM.

Chapter 4

Workload Prediction: Initialisation

The previous chapter presented the aspects considered to build our rule-base incrementally, walking through the various cycles of the DSRM approach. This chapter presents the steps followed to build the initial patient model that drives the subsequent prediction process presented in the next chapter.

Our adaptable evidence-based workload prediction model (AE-WPM) operates in two phases, namely initialisation and prediction. In this chapter, first I discuss, in the context of initialisation of the model, the various aspects of this workload problem. I then discuss the challenges, assumptions, and definitions; and further, explain the rules (decision tables) developed to prepare data for the subsequent prediction phase. Finally, I discuss the validation of a few assumptions and rules developed for preprocessing and preparing the patient model for the simulation.

4.1 The workload management challenge - a recap

A primary health care system provides general care including acute needs, LTC management needs and immunisations (WHO, 1978). In order to access these services, a patient books an appointment with a GP at the PHC centre (Montague, 2014). During the consultation, the GP goes through the patient's current and previous medical history, then accordingly, lab tests may be requested, medications may be prescribed and a follow-up[1] may also be suggested. Accordingly, the next appointment may be booked by or scheduled for the patient.

However, the growing demand for care due to patients with LTCs is posing a ma-

[1]The terms follow-up, next appointment and recalls are used interchangeably in this work to refer to the subsequent appointments by a patient.

jor challenge on the health care domain. Consequently, it is possible that a patient might not be able to make an appointment on the date as advised by the GP due the unavailability of GPs. Moreover, (multi-morbid) LTC management requires timely recalls to prevent and detect associated complexities early, and to better manage the LTCs present in a patient. Therefore, much emphasis is placed on the need to manage workload to provide better care (Bodenheimer et al., 2009). Hence in this work, we make an attempt to predict the upcoming workload as a volume of the number of appointments from the LTC patients following their plan-of-care.

4.2 The workload prediction process

The workload prediction process operates in two steps, namely initialisation and simulation. Figure 4.1 gives an overview of this prediction process. As depicted in the figure, medical practice refers to care pathways to meet the LTC management needs of their patients. These care pathways are fed (as rules) into the simulation phase. The original patient dataset is missing information required for our work (see Section 4.4.1) and therefore needs to undergo a pre-processing step (shown as an arrow marked A) to enrich the data with inferred patient information needed for our work. The patient records are then filtered to retain information up until the simulation start date[2].

Along with the care pathways, this initial patient model is an input to the simulation phase. The simulation then builds care plans (the arrow B denotes the iterative cyclic approach presented in Section 3.3 that extends the rule-base for simulation) for each individual. Based on these individual care plans, our simulation predicts the recalls for these LTC patients, which is then aggregated as the predicted workload.

Our simulation walks through time (from a given simulation start date to a simulation end date). Therefore, the initial patient model has the patient details such as individual patient problem classes and their corresponding LTC appointment dates (derived from the pre-processing steps) prior to the simulation start date.

Based on the individual problem classes applicable to a patient, for each patient our rules aggregate their care plans to decide the patient's recall period as the most frequent recall period for the patient's various conditions. Then, based on their most recent LTC visit date prior to the simulation start date, and the most frequent recall period for a patient, the subsequent recall decision is made for each patient. This next

[2]We have real visit data for the year 2014. We simulate workload for 2014 so that we can compare the simulation results with the actual behaviour

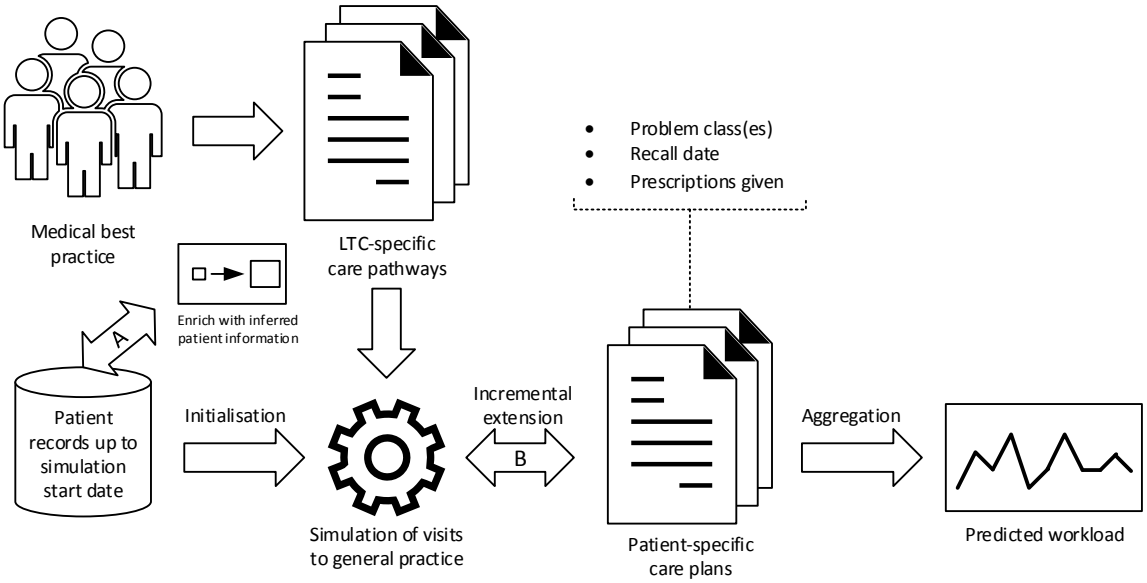


Figure 4.1: Overview of the workload prediction process modified[3] from Devananda et al. (2017).

appointment is calculated as their visit date plus the most frequent recall period (in months) for the patient. This recall is recorded as a fact in the rule engine's working memory. This continues for the simulation period. Finally, after the simulation ends, these recorded recalls are aggregated to a weekly basis to give the predicted workload.

As discussed in Section 2.2.5, the use of rule-based models in the health care domain is widely accepted. Knowledge-based clinical decision support systems use an "If-Then" rule base applied onto the patient data to propose care actions for the patients (Alther and Reddy, 2015). Shiffman (1997) and Minutolo et al. (2017) also highlight the benefits of using a rule-based system in medical settings, which includes the following.

1. We can extend the rule-base by adding new rules; this feature helps us to incrementally build and evaluate our workload prediction model through the DSRM cycles presented in Chapter 3.

2. We have the flexibility to choose different rules based on the facts expressing the current context; this feature help us to build the plan-of-care from care plans.

3. It enables dynamic handling of changes through rules being re-activated based on changes to facts during the execution of rules.

[3]Updated problem 'codes' to problem 'class(es)'.

The above two (2 & 3) features are essential to explore our what-if scenarios. Therefore, we use a rule-based approach in this work.

The following section presents the rule engine Drools and our rules in Drools rule formats.

4.3 The Drools rule engine and our rules

This work uses the Drools rule engine (Salatino et al., 2016). Drools is a Business Logic integration Platform (BLiP). The use of Drools has two main stages:

1. Authoring: During authoring, we create rules in .drl format for Drools. These files contain the rule definitions in a declarative form. Drools also supports a decision table format, which gets converted (internally) to .drl format at run time.

2. Runtime: At runtime, we create the working memory, which is a representation of the input data as 'facts'. The rules are then applied repeatedly to the incoming data, until no more rules are triggered.

A Drools rule (hereafter referred to as a D-rule) has a condition and an action part. Every D-rule is identified by a rule name. Conceptually, at each point all rules' conditions are evaluated against the state of working memory. Rules with true conditions are activated. Once all the rules are evaluated, the activated rules get fired and subsequently, some facts are modified, inserted or deleted.

One problem that can arise from rule-based systems is potential conflict between multiple rules that are triggered in the same context. In our work, the number of rules was small and the rules are designed to avoid conflict.

By default the choice of the rule to fire next is random; the *salience* feature can be used to control the logical sequence of the rules' evaluation and execution. We have used salience extensively in our pre-processing and in simulating what-if scenarios. For example, when one visit among multiple visits of a patient is shifted to nurses, the logic applied is to schedule all the visits to GPs initially and on the simulation end day, one random visit among various visits of the patients is shifted to nurses. This should happen only after the visits for that day are scheduled to GPs. So, salience is used to order the execution of activated rules.

By default whenever the state of working memory changes, i.e., when a fact gets inserted, modified or deleted, all rules are (re)evaluated and consequently the set of

activated rules may change. The *no-loop* feature of a rule can be used to avoid re-evaluation of the rule. Similarly, there are other features in Drools that provide flexibility in using Drools. Salatino et al. (2016) give an exhaustive list of features available in Drools.

4.3.1 The Drools objects and facts

Every fact in the working memory is associated with an object in the Java programming language. Objects comprise attributes or members. Facts associate values to these attributes. In our case, we define object classes based on our data tables (see Figure 4.2). These classes include `Patient`, `Classification`, `LatestLabResult`, and `Recall` (see Appendix A.2 for a complete list of classes and related attributes used in our work).

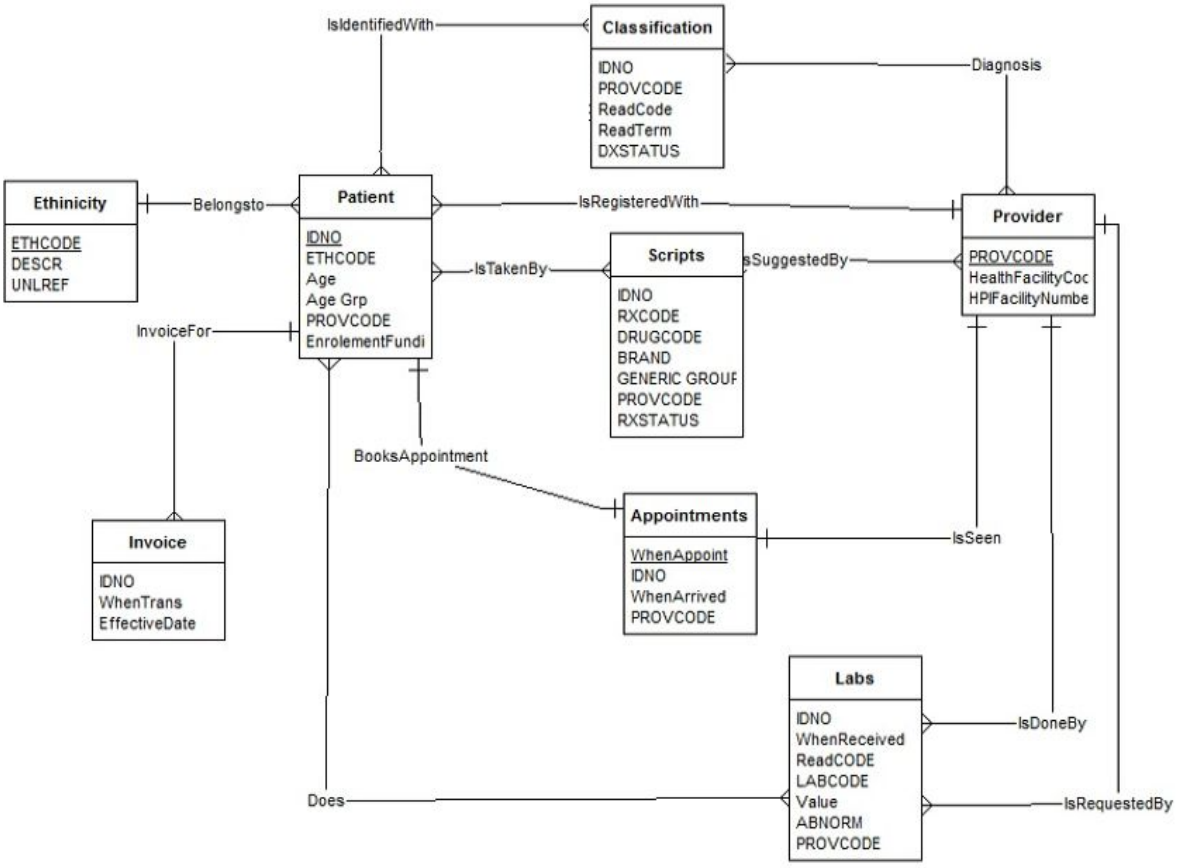


Figure 4.2: A partial entity-relationship diagram.

A Patient object has `patientId`, `age`, `ethnicity`, `prov`[4] and other attributes. A

[4]prov captures the provider role code

Classification has `patientId`, `dateOfClassification`, `Readcode` and so on. Likewise, a Recall has `patientId`, `recallDate` and so on. In order to specify a condition, we use a class name and corresponding criteria on the attribute to be met. For example to check if a patient is above age 50, we write:

```
Patient(age>50)
```

Having given a brief idea on how facts are represented and accessed, the following section will discuss the Drools rule formats.

4.3.2 Drools rules formats

Drools supports two different styles of rule representations: the .drl format and the decision table format. In this work, while preprocessing of data for initialisation uses decision tables, the prediction process uses the .drl rule format. Before diving into a more detailed discussion of rule formats, let us consider an example of the diabetic risk score calculation for a patient as taken from NZGG (2011). Table 4.1 shows a tabular form of variables and conditions that is used to calculate the risk score for diabetes (`$drs`). The column *Variable* indicates various parameters coded using a corresponding *Code*. If a patient's data values of the corresponding coded parameter match the *Criteria*, then the risk factor of diabetes for the patient increases by the *Score* given in the respective row. For example, the first row of the table reads: If a patient's Hba1c in mmol/mol coded using Code 44TB is greater than 55, then add 1 to the risk score of the patient.

Table 4.1: Diabetic risk score calculation factors taken from NZGG (2011).

Variable	Code	Criteria	Score
Hba1c mmol/mol	44TB	>55	+1
BP Systolic and diastolic mmHg	2469 246A	> 130 and > 80	+1
ACR mg/mmol	46TD	> 3	+1
eGFR CKD-EPI ml/min/1.73^2	44J3 (serum Creatinine)	<60	+1
Triglycerides and Tot Chol.	44Q 44P	>= 1.7 and >= 4.0	+1
Smoking Status*		Smoker	+1
Ethnicity	(code 20-39 or 41-43)	Maori or Pacific Island or South Asian	+1

.drl format

Traditional rules with a "when-then-end" format are stored as *.drl* rules. The first row of Table 4.1 is given as a rule in Listing 4.1, which reads: If a patient's Hba1c

coded using Code 44TB is greater than 55, then add 1 to the risk score of the patient. In Listing 4.1, we use $ to indicate it is a variable of that class. For example, line 4 indicates that $p is a variable of type `Patient` class. So, all the attributes of the `Patient` class are available to $p.

The `first_row_of_drs_calculation_table` (line 1) is the rule name. The `when` keyword denotes start of conditions of the rule. Every condition within the `when` block are "AND"ed together. The `then` keyword marks the action part of the rule. The statements in the `then` block are executed sequentially. The `end` marks the end of the rule.

In this example, we need to update those patients (`$p`) whose lab results for `44TB` value is greater than `55` (line 5), **and** have a diabetic risk score `$drs` associated with them (line 6), **then** add one to their risk score ($drs) (line 9).

Listing 4.1: .drl rule for first row in Table 4.1.

```
1 rule "first_row_of_drs_calculation_table"
2 when
3 //conditions are "AND"ed together in this block.
4   $p:Patient()
5   $lab:LatestLabResult(patientId == $p.patientId,readCode=="44
     TB", labValue>55)
6   $drs:DiabeticRiskScore (patientId == $p.patientId)
7 then
8 //The statements in the then block are executed sequentially.
9   $drs.riskScore = $drs.riskScore + 1;
10  modify{$drs};
11 end
```

Subsequent rows are appended, similarly, as individual rules to the rule above.
Another way to represent rules in Drools is using decision tables.

Decision table format

A decision table is a matrix of rows and columns, where each row of a decision table forms a rule. Columns represent the criteria of the conditions. Each such decision rule states what procedure to follow when certain conditions exist (Salatino et al., 2016; Senn, 1989).

In Drools, decision tables can be represented using a spreadsheet in either `.xls` or

`.csv` format[5]. Figure 4.3 implements Table 4.1 as a Drools decision table. Internally each row of the decision table is converted into a D-rule. Listing 4.2 presents D-rules for rows 38 and 39 of the decision table in Figure 4.3.

Listing 4.2: Example of rows 38 and 39 in the decision table in Figure 4.3 converted internally to .drl rules.

```
1 // rule values at C38, header at C33
2 rule "DiabeticsRiskScoreCalculator_38"
3 when
4   $classifi:Classification(problemCode == "DM_COD")
5   $latestlabresult:LatestLabResult(patientId == $classifi.
    patientId,labCode=="44TB",labValue >55)
6   $drs:DiabeticRiskScore(patientId==$classifi.patientId,
    riskScore <4)
7 then
8   $drs.riskScore = $drs.riskScore+1; update($drs);
9 end
10 // rule values at C39, header at C33
11 rule "DiabeticsRiskScoreCalculator_39"
12 when
13   $classifi:Classification(problemCode == "DM_COD")
14   $latestlabresult:LatestLabResult(patientId == $classifi.
    patientId,labCode=="2469",labValue>130)
15   $otherlabresult:LatestLabResult(patientId == $classifi.
    patientId,labCode=="246A",labValue>80)
16   $drs:DiabeticRiskScore(patientId==$classifi.patientId,
    riskScore<4)
17 then
18   $drs.riskScore = $drs.riskScore+1; update($drs);
19 end
```

Every decision table has a global configuration section (rows 32 to 37 in the figure) and a rule declaration section (rows 38 and below).

The global configuration section has keywords to indicate the rule package, the start of a rule table, the table name to be used to auto generate rule names, to specify if the cell contents belong to the left hand side (LHS) i.e., the condition or the right hand side (RHS) i.e., the action of a rule, and the classes of object to which these

[5]I use .xls format in this work.

	B	C	D	E	F	G	H	I
32		**RuleSet**	predictionPack					
33		**RuleTable DiabeticsRiskScoreCalculator**						
34		CONDITION	CONDITION	CONDITION	CONDITION	CONDITION	CONDITION	ACTION
35		$classifi:Classification	$latestlabresult:LatestLabResult	$otherlabresult:LatestLabResult	$otherclass:Classification	$p:Patient	$drs: DiabeticRiskScore	
36		problemCode == "$param"	patientId == $classifi.patientId, labCode=="$1", labValue $2	patientId == $classifi.patientId, labCode =="$1", labValue $2	patientId == $classifi.patientId, problemCode == "$param"	ethinrange(ethCode), patientId==$param	patientId==$param,riskScore <4	$drs.riskScore = $param; update($drs);
37	**Recall Decisio**	**Check problemCode**						**decide recall**
38	**DM_COD**		**44TB,>55**				$classifi.patientId	
39	**DM_COD**		**2469,>130**	**246A,>80**			$classifi.patientId	
40	**DM_COD**		**46TD,>3**				$classifi.patientId	
41	**DM_COD**	**DM_COD**	**44J3,<60**				$classifi.patientId	**$drs.riskScore+1**
42	**DM_COD**		**44Q,>=1.7**	**44P, >=4.0**			$classifi.patientId	
43						**$classifi.patientId**	$classifi.patientId	
44	**DM_COD**				CSMOK_COD		$classifi.patientId	

Figure 4.3: Drools decision table to calculate diabetic risk score.

conditions or actions apply. The rule declaration section specifies the criteria on the objects specified in the global configuration section. Thus, the global configuration section depicts how to map rows of the rule declaration section to the corresponding "rule-when-then-end" sections of a D-rule.

The `RuleSet` keyword (in row 32 in the figure) indicates the name to be used in the rule package that will encompass all the rules. The `RuleTable` keyword (on row 33) marks the start of this decision table. Any text following the `RuleTable` keyword (DiabeticsRiskScoreCalculator, here) on the same row is considered as a part of the auto generated *rule-name* for each row of that table. The row numbers are appended to guarantee unique rule names. For instance, the rule name corresponding to row 38 will be "DiabeticsRiskScoreCalculator_38" (see the first rule in Listing 4.2).

The keywords `CONDITION` and `ACTION` indicate that the data in the columns below are for the `when` and the `then` parts of a rule, respectively. For a given row, the column of the RuleTable entry indicates the column in which the rules start; columns to the left are ignored. Therefore, entries in columns A and B are ignored in this case. The rule is built from rows 38 onwards and conditions for the rules are decided from columns C to H.

In Listing 4.2, you can infer that the rules are built horizontally, based on conditions built vertically from the columns. In other words, the conditions are "AND-ed" horizontally to form a rule and are "OR-ed" vertically to append new rules to the rule-base. Also, as seen in Listing 4.2, the values in the cells below the condition classes become constraints on that object type. For instance, in the figure, consider column C. The cell entry `$classifi:Classification` indicates that the condition must be applied to facts of type Classification. They can use the notation `$param` as a place holder to indicate where data from the cells below should be interpolated. In the figure, the criterion for this column is specified as `problemCode == $param`. This is replaced with `problemCode == "DM_COD"` as the first condition in rules, when the Column C cells in these rules have the value "DM_COD". For multiple insertions, we use `$1`, `$2`, etc., indicating positions of parameters in a comma-separated list in a cell below.

A separate column is used to define each subsequent condition for a rule, as shown in columns D to H. It is possible that certain conditions apply only to some rules in the set. For instance, the second rule row of Table 4.1 sets criteria for both `2469` and `246A`. In order to specify this (additional) "AND" condition, in the decision table we use `CONDITION` column D. As seen in the second rule in the listing, there are two conditions (`DiabeticsRiskScoreCalculator_39`). When the conditions don't apply

to a rule, the corresponding cell is left empty (e.g., in row 38, columns E, F and G are left empty and indicate those conditions do not apply to this rule).

An `ACTION` defines the rule action. Here, the rule action is to update `$drs` of the patient by adding 1 to its `riskScore` attribute (see column I). Rule rows are read until there is a blank row.

Apart from these basic features, each rule can also be assigned a different name using a `NAME` column type. We can also use a column to indicate the `PRIORITY` of a rule. It has to be noted that multiple `RuleTable`s can exist in a sheet. More features and implementation details are given by Salatino et al. (2016).

Having introduced a broad idea of how the rule base is represented, we describe the initial dataset, the rules for pre-processing this initial dataset, and the validation results of the developed rule base.

4.4 Initial dataset

The dataset[6] includes a patient Register which records details of 29 974 patients, an Appointments table, a Scripts table to store the medication prescription details, a Classifications table, a Transactions table, an Invoice table, an Immunisations table, a table Ethnicity that records details of various ethnic codes used in the dataset and a Labs table.

New Zealand has a public-share-funded health care system (Ministry of Health NZ, 2016e), and our collaborator in this work was interested in the workload from their funded patients. Hence, we consider only the 9 895 publicly[7] funded patients and only a subset of these funded patients (as described in the following section) belong to our cohort. For instance, to be included in our cohort a patient must be an LTC patient and must have made an LTC appointment within a defined period.

The clinical support system of our collaborator uses MedTech32 (MedTech Limited, 2013), and so queries to define the cohort use tags specific to the MedTech32 data formats. For instance, funded patients are recorded with an 'F' for enrolmentfundingcode attribute of a patient in the patient register. (See Appendix A.11 for a partial descrip-

[6]As mentioned in Chapter 1, the data was anonymised before being supplied and the University of Otago Human Ethics Committee (Health) approved the use of this anonymised data in our study.

[7]In a New Zealand context, there is no notion of "privately funded" patients. There are either publicly funded or non-funded patients. To be eligible for public funding, the patients should meet the eligibility criteria as required by the Minister of Health under section 32 of the New Zealand Public Health and Disability Services Act 2000 (Ministry of Health NZ, 2016e).

tion of the mapping from the underlying MedTech32 database to the tables used in this study).

I take this opportunity to remind the reader that once the patient data has been preprocessed, it represents a generic patient with similar LTC conditions rather than the actual patient. Our simulation model does not simulate patients leaving or joining the practice. We also do not simulate new LTCs being diagnosed or the LTC deterioration in a patient, and so do not simulate patient outcomes.

4.4.1 Shortcomings of the dataset and compensating with inferences

The dataset captures individual patient data. However, the dataset is missing certain information that is needed, and has more granular data than required for this study. Hence, the data needed to undergo some preprocessing, in order to infer our required information.

For instance, consider a patient A. The patient register does not explicitly identify A as an LTC patient. It is not evident from the appointments by A whether they were related to LTCs or not. Apart from the most recent script of an LTC-related medication for A being tagged with a 'T', no information explicitly indicates that the medication was prescribed to manage a specific LTC in A. Similarly, in other tables, no data or information explicitly captures whether an event, action or data captured in these tables relate to LTC(s) in a patient. Thus, one major shortcoming found from analysing this health data was that it does not explicitly represent the LTC status of an entity in these tables. Hence, we applied some general assumptions to identify LTC medications and the currently active medications in a patient which, in turn, were used to identify LTC-related appointments (hereafter referred to as LTC appointments) of a patient. These assumptions and definitions were fine-tuned over many iterative cycles of evaluation via feedback from our PCA after studying statistics on the resulting patient database, simulated workload, and examining generated appointment traces for individual patients.

The information inference logic adopted for this study is primarily advised by our PCA. Here, his expertise as a clinical manager and as a GP play a major role. So, note that these definitions and assumptions could vary from practice to practice and care provider role to role.

Identifying LTC appointments

In order to start the patient visits simulation, the model needs to know when the last LTC appointment was made by each patient of the cohort, given the simulation runs for a given time-frame specified by a *start_date* and an *end_date*. As the appointments table lacked explicit representation of the reason (such as LTC, acute, exacerbation of an LTC and so on) for an appointment, one of the major challenges was to identify the LTC appointments by these patients.

Generally, administrative and billing systems capture a limited but consistent set of data about a patient and the service provided (Steinwachs and Hughes, 2008). In our case, the transactions table records which role is responsible for a transaction in the `sercode` of the `transaction` table. Hence, we looked for those appointments whose transactions were related to a GP or a nurse. In addition to that, we also assume that LTC medications are prescribed every 3 months, and patients tend to visit a practice when their medications are due to run out. Although usually patients come earlier than their medications run out, there are cases when they come later. Hence, in this work, we define an LTC appointment as any appointment that

- involves a consultation with a GP or nurse (`sercode` of the `transaction` is encoded with a 'C' or an 'NC'); and,
- has a percentage (see below) of active long-term medications prescribed within a three-week window (two weeks before and one week after) of the appointment date of a patient.

During an appointment related to non-LTC related needs, only those LTCs (if any) that may be affected by medications to that non-LTC will be addressed. However, during an appointment related to an LTC(s), usually all the LTCs and their medication requirements are discussed and medications are prescribed accordingly for the patient. Therefore, we consider the percentage of active long-term medications prescribed.

The reason for considering a three-week window is because the medications may be prescribed before or after an appointment. For instance, some patients come earlier than their medications run out. It is also possible that during an appointment the GP would suggest some lab tests and later during the week, the GP may prescribe (or change) medications based on those lab test results. Therefore, we consider a three-week window (two weeks before and one week after) of the appointment date of a patient.

Identifying LTC prescriptions

The `scripts` table records details including the patient (`patientId`), the drug code, the generic name of the medicine and the brand of the medicine. However, it does not capture the reason for prescribing a specific medication for a patient. Moreover, medications belonging to the same generic medications may differ in their brand and dosage. The generic names were not null, however, there were incomplete[8] generic names in the table. For example, the encoding of `PARACETAMOL` (the generic name of the drug) prescribed varied based on the drug code (such as `MT-2902` and `MT-511`) and the brand name (such as `PARACETAMOL`, `PAMOL`, and `PANADOL`). Hence, we considered the first six letters of the generic name of the medications to decide the count of medications prescribed to a patient. For the above example, we use `PARACE` for all medications prescribed with a generic name `PARACETAMOL`.

The calculation of the number of medications prescribed on a date of prescription for a patient involved three steps. For each patient:

Step 1 : We made a list of the T-tagged medications prescribed historically for that patient. Please recall that the MedTech32 database uses a data tag 'T' in the most recent prescription of a medication to indicate that it was prescribed for an LTC in the patient (although it does not explicitly say which LTC(s)). Then,

Step 2 : We counted how many T-tagged medications were *active* on a date of appointment of the patient. An *active* medication is a medicine prescribed for a patient within four months of the date of appointment of the patient. This is because LTC medication prescriptions may be valid for up to four months. Next,

Step 3 : We compared the number of total active medications the patient is on and the number of medications prescribed within the three week window (as mentioned above) of the date of appointment of the patient[9]. Then, we defined a patient as prescribed with an LTC prescription if the patient was:

- not on any active LTC medicine but an LTC medication was prescribed within the three week window (this is probably due to a new diagnosis or a change of medication for an LTC in the patient); or
- on one active LTC medicine and there was one (or more) LTC medication prescribed within the three week window; this may be due to one or more

[8]Only the first few characters of the name were recorded.

[9]These rules were validated by our PCA.

factors such as other LTCs being diagnosed or prevention medications prescribed or LTC severity having changed in the patient; or

- on two active LTC medicines and at least one LTC medicine was prescribed within the three week window i.e., at least one LTC in the patient has to be managed through medication; or
- on three active LTC medicines and at least two LTC medicines were prescribed within the three week window; or
- on more than three active medicines and at least half of the number (rounded down) of the active medicines were prescribed within the three week window for the patient.

Yet another shortcoming was related to care delivery decisions. The data captures neither the reasons for the care actions nor the care plan identifiers or the care pathways applied to a patient. Hence, a few decision tables were developed to infer which care plans of a care pathway were applied to a patient. Furthermore, we also needed to classify LTC severity in the patients. The following sections discuss these decision tables, which form a rule base to filter and refine data according to our requirements. The resulting data set was then used to seed to execute our workload simulation model. In New Zealand, the Ministry of Health makes effort to record patient demographics specifically Māori and Pacific Islanders (M&P) and their issues. However, the clinical guidelines applied or referred to manage a specific LTC are those developed in the US or the UK. According to our PCA, there are no clinical guidelines to manage an LTC specifically in a M&P group, although there are health strategies (Ministry of Health NZ, 2000) that aim to improve access to care, low cost access and so on to address health issues of the M&P group. So, for M&P group, the practitioners continue to refer to the standard guidelines available.

4.5 Data pre-processing for simulation

A partial entity-relationship (ER) diagram of the dataset was introduced in Chapter 3. Here, I will discuss the preprocessing steps required to prepare data (i.e., our initial patient model) for seeding the simulation of patient recalls.

4.5.1 Decision table: to map Read codes to problem class(es)

As noted earlier, in this study we do not simulate patient outcomes. Hence, the information we needed to know was the LTC in a patient (the general name rather than the specific clinical encoding of the LTC), and the corresponding recall period for the patient. However, clinical coding is used extensively within the PHC system (Alther and Reddy, 2015). In this practice dataset, the clinical coding system was Read V2. Clinicians are encouraged to code as far down the coding hierarchy as possible to increase the accuracy and specificity of disease coding (Mabotuwana and Warren, 2010). So, there are multiple Read codes associated with each LTC, e.g., any Read code starting with G2, such as G20 and G27, is used to represent hypertension and more information related to hypertension in the patient. Since these Read codes capture more details than we need, we mapped these Read codes to problem classes. Problem classes are a more generic name for the LTCs; e.g., following on from the previous example, both the Read codes G20 and G27 will be mapped to a problem class hypertension (which we encode as `HYP_COD`[10]).

In this work, we focus on Read codes that identify one of the 12 LTCs listed below. The label for each LTC in this work is given in parentheses.

1. Chronic Kidney Disease (CKD)
2. Hypertension (HYP_COD)
3. Diabetes diagnostic codes (DM[11]_COD)
4. Coronary Heart Disease diagnosis codes (heartdisease)
5. Stroke diagnosis codes (stroke)
6. TIA diagnosis Codes (tia)
7. PAD diagnostic codes (pad)
8. Atrial Fibrillation (atrial)
9. Congestive Heart Failure (heartfailure)
10. COPD Diagnostic codes (COPD)

[10]COD stands for condition.
[11]Diabetes Mellitus

11. Asthma Diagnostic codes (asthma)

12. Psychosis, schizophrenia & bipolar affective disease codes (psychosis)

Figure 4.5 presents a partial Drools decision table that maps Read codes to problem class(es) (please refer to Appendix A.3 for the complete decision table).

4.5.2 Decision table: to identify the CKD stage in a patient

As already mentioned in Section 3.2.2, the severity or presence of Chronic Kidney Disease (CKD) in a patient is not recorded by Read codes, instead the CKD stage in a patient is determined by two factors, namely the Gfactor (eGFR) and ACR[12] stage of the patient. With the Gfactor and ACR values, the recall is decided according to the recall plan matrix given in Figure 4.4. It defines CKD recall periods based on the two factors. It also presents the severity of the CKD stage in a patient. Here, note that green means no CKD (hence no recall), yellow means low severity (12 month recall), orange means medium severity (six month recall) and red means most severe (three month recall).

Follow up time period in Months for Chronic Kidney Disease					
			Persistent albuminuria categories Description and Range		
		Null	A1	A2	A3
GFR categories (ml/min/ 1.73m2) Description and range	G1	No CKD	No CKD	12/12	6/12
	G2	No CKD	No CKD	12/12	6/12
	G3a	12/12	12/12	6/12	3/12
	G3b	6/12	6/12	3/12	3/12
	G4	3/12	3/12	3/12	2/12
	G5	3/12	3/12	3/12	2/12

Figure 4.4: CKD stage recall plan.

We use the CKD-EPI equation[13], which is a standard one for calculating the G-factor of a patient (Levey et al., 2009). Because the parameter values depend on patient details (see Table 4.2), the decision table in Figure 4.6 selects the right parameters to

[12] Albumin-Creatinine Ratio

[13] $eGFR = factor$*(((*serum creatinine in* $\mu mol/L$)*0.0113/*denominator*)^*exponent*)*0.933^*age*

98		RuleSet	predictionPack		
99		RuleTable Assigning ProblemCode			
100		CONDITION	CONDITION	ACTION	ACTION
101		classifi : Classification			
102		latestReadCode matches "$param",problemCode==null	latestReadCode not in ($param)	classifi.problemCode = "$param";	System.out.println(classifi.patientId +" " + classifi.problemCode);System.out.println("$param"); update(classifi);
103	**Problem Types**	**ReadCodes**	**Not ReadCode**	**Problem Code**	**Print Success**
104	**HyperTensionCode**	**G2**		**HYP_COD**	**Problem Code hypertension is set**
105		**G20.***			
106		**Gyu2**			
107		**Gyu20**			
108	**Heart Disease**	**Gyu3.***	**"Gyu31"**	**heartdisease**	**Problem Code heart disease is set**
109	**Stroke**	**G61.***	**"G617"**	**stroke**	**Problem Code stroke is set**
110		**G64.***			
111		**G66.***	**"G669"**		
112		**G6760**			
113		**G6W**			
114		**G6X**			

Figure 4.5: Mapping Read codes to problem classes: a partial Drools decision table.

be used in the equation. The AssignFactor function in Listing 4.3 then takes these corresponding parameter values to calculate the Gfactor. Our data (as part of the anonymisation) does not capture the date of birth but has the age of the patient in 2015. Therefore, we calculate `newage` to be used in the equation. The data also has records that incorrectly recorded the serum creatinine values, hence we needed to correct it with a multiplication factor 1000 (when the value is less than 2).

Listing 4.3: User defined function to assign Gfactor based on various parameters.

```
1 function int AssignGFactor(int factor, double ser_creat, double
      denom, double expo, Date labreceiveddate,int age)
2 {
3     int newage = age - (2015 - labreceiveddate.toLocalDate().
    getYear());
4     if(ser_creat < 2)
5        ser_creat = ser_creat * 1000;
6     double eGFR = factor * Math.pow(((ser_creat*0.0113)/denom),
    expo) * Math.pow(0.993,newage);
7     int eGFRValue = (int) eGFR; //floor value of the double
    needed
8     return (eGFRValue);
9 }
```

Table 4.2: CKD-EPI parameter-values

Ethnicity	**Gender**	**Serum Creatinine**	**Factor**	**Denominator**	**Exponent**
Black	Female	<=62	166	0.7	-0.329
Black	Female	>62	166	0.7	-1.209
Black	Male	<=80	163	0.9	-0.411
Non Black	Female	<=62	144	0.7	-0.329
Non Black	Female	>62	144	0.7	-1.209
Non Black	Male	<=80	141	0.9	-0.411
Non Black	Male	>80	141	0.9	-1.209

	CONDITION	CONDITION	CONDITION	CONDITION	CONDITION	CONDITION	ACTION	ACTION	ACTION	ACTION
	$p:Patient			$latestlabresult: LatestLabResult						
	ethCode == $param	ethCode != $param	gender == '$param'	patientId == $p.patientId, labCode == "$1"	(int)labValue <=$param	(int)labValue > $param	System.out.println("In the rule Gfactor Calculation");int $factor = $param;	double $denom = $param;	double $expo = $param;	CKDLabResult $newlabresult = new CKDLabResult(); $newlabresult.patientId = $latestlabresult.patientId; $newlabresult.receivedDate= $latestlabresult.receivedDate;$newlabresult.labCode = "eGFR"; $newlabresult.labValue = $param; insert($newlabresult);retract($latestlabresult);
Problem Types	**Check Ethnicity Black**	**Check Ethnicity NonBlack**	**Check Gender**	**Lab Code for Serum Creatinine**	**<= condition**	**> condition**	**Factor**	**Denominator**	**Exponent**	**Assign G Factor and update the same**
BLACK Female	**54**		**F**	**44J3**	**62**		**166**	**0.7**	**-0.329**	**AssignGFactor($factor,Double.valueOf($latestlabresult.labValue),$denom,$expo,$latestlabresult.receivedDate,$p.age)**
						62			**-1.209**	
BLACK Male			**M**		**80**		**163**	**0.9**	**-0.411**	
						80			**-1.209**	
Non-BLACK Female		**54**	**F**		**62**		**144**	**0.7**	**-0.329**	
						62			**-1.209**	
Non-BLAck Male			**M**		**80**		**141**	**0.9**	**-0.411**	
						80			**-1.209**	

Figure 4.6: eGFR parameter assign table.

CONDITION	CONDITION	CONDITION	CONDITION	CONDITION	ACTION	ACTION
$latestlabresult: LatestLabResult						
labCode == "$1"	(int)labValue == $param	(int) labValue != 0 && (int) labValue < $param	(int) labValue >=$1 && (int) labValue <=$2	(int) labValue > $1	int ACRStage = $param;	CKDLabResult $newlabresult = new CKDLabResult(); $newlabresult.patientId = $latestlabresult.patientId; $newlabresult.labCode = "ACR"; $newlabresult.labValue = $param; $newlabresult.receivedDate= $latestlabresult.receivedDate; retract($latestlabresult);
Lab Code for Serum Creatinine	**0 condition**		**>= and <= condition**	**> condition**	**ACR stage**	**AssignACRand update the same**
46TD	**0**				**1**	**ACRStage**
		3			**1**	
			3,30		**2**	
				30	**3**	

Figure 4.7: ACR calculation in a decision table.

Similarly, the decision table in Figure 4.7 assigns the ACR-stage in a patient based on values of the ACR lab test results (ACR lab tests are Read coded as 46TD; and see Table 4.3). Table 4.4 maps the results of eGFR and ACR stage to decide the CKD stage in a patient. Therefore, the results of these two decision tables are inserted into the working memory which then, together, determine the CKD-stage as given in Figure 4.8. Please refer to Appendix A.4 for the complete CKD care pathway used in this work.

Table 4.3: ACR stage assign criteria.

ACR	**Stage of Protein Loss**
absent (0)	A1[14]
<3	A1
>=3 and <=30	A2
>30	A3

Table 4.4: Mapping latest eGFR values from Figure 4.6 and ACR from Figure 4.7 to CKD Stage.

eGFR	**ACR**	**Stage of CKD**
<=14	A3	G5 A3
<=14	A2	G5 A2
<=14	A1	G5 A1
<=14	Null	G5
>=15 and <=29	A3	G4 A3
>=15 and <=29	A2	G4 A2
>=15 and <=29	A1	G4 A1
>=15 and <=29	Null	G4
>=30 and <=44	A3	G3b A3
>=30 and <=44	A2	G3b A2
>=30 and <=44	A1	G3b A1
>=30 and <=44	Null	G3b
>=45 and <=59	A3	G3a A3
>=45 and <=59	A2	G3a A2

[14]As per the care plan, it should assign Null. We use A1 as per the best practice guidelines suggested by our PCA.

>=45 and <=59	A1	G3a A1
>=45 and <=59	Null	G3a
>=60 and <=90	A3	G2 A3
>=60 and <=90	A2	G2 A2
>=60 and <=90	A1	No CKD
>=60 and <=90	Null	No CKD
>=90	A3	G1A3
>=90	A2	G1A2
>=90	A1	No CKD
>=90	Null	No CKD

4.5.3 Decision table: to calculate the diabetic risk score of a patient

Diabetes in a patient can become more complicated due to prolonged glycosylated haemoglobin level in one's blood stream (Papatheodorou et al., 2015). This complexity is depicted by a *risk score* of diabetes in a patient. As depicted in the care pathway for diabetes, (please refer to Appendix A.5 for the complete diabetes care pathway used in this work), patients are classified based on their risk score (from 0 to 3 where 3 is high risk). However, if the patient has been diagnosed with heart disease, stroke, TIA or PAD, then they are categorised as a very high risk patient. We use risk score 4 to depict these very high risk patients (refer to the last four rows in Figure 4.9). The other rows, based on laboratory results, add one to the risk score of a patient.

As depicted in the care pathway, the smoking status of a patient also plays an important part in deciding the risk score of the patient. The smoking status of the patients are current smokers, ex-smokers and non-smokers. I would like to highlight the difference between the most accepted NICE[15] guidelines and the best practice guidelines for diabetes management followed in this work. The NICE guideline for diabetes takes into account whether the patient has ever smoked, how long this smoking status was unchanged and the current smoking status to determine the risk factor for a patient (NICE, 1999). On the other hand, the best practice guidelines for diabetes applied in this study are those of the NZGG (2011) and consider only the current smoking status of the patient as a factor to calculate the diabetic risk score. According to the

[15]The National Institute for Health and Care Excellence, UK.

RuleTable Assign CKD Stage				
CONDITION	CONDITION	CONDITION	CONDITION	ACTION
$ACRlabresult: CKDLabResult	$eGFRlabresult: CKDLabResult			
labCode == "ACR" , (int) labValue == $param	labCode == "eGFR", patientId == $ACRlabresult.patientId,labValue <= $1	labCode == "eGFR", patientId == $ACRlabresult.patientId, labValue >= $1 && labValue <=$2	labCode == "eGFR", patientId == $ACRlabresult.patientId,(int) labValue >= $1	System.out.println("Rule fired for CKD stage for idno " + $eGFRlabresult.patientId); Classification $class = new Classification();String CKDStage = "$param"; $class.patientId = $eGFRlabresult.patientId; $class.problemCode = "CKD"; $class.latestReadCode = "$param"; $class.dateOfClassification=$eGFRlabresult.receivedDate; insert($class);
Lab Code for ACR	**check the eGFR avlue**	**check the eGFR avlue**	**check the eGFR avlue**	
3	**14**			**G5 A3**
2				**G5 A2**
1				**G5 A1**
3		**15,29**		**G4 A3**
2				**G4 A2**
1				**G4 A1**
3		**30,44**		**G3b A3**
2				**G3b A2**
1				**G3b A1**
3		**45,59**		**G3a A3**
2				**G3a A2**
1				**G3a A1**
3		**60,89**		**G2 A3**
2				**G2 A2**
1				**No CKD**
3			**90**	**G1 A3**
2				**G1 A2**
1				**No CKD**

Figure 4.8: eGFR and ACR used to determine CKD stage.

	RuleSet	predictionPack						
	RuleTable DiabeticsRiskScoreCalculator							
	CONDITION	CONDITION	CONDITION	CONDITION	CONDITION	CONDITION	ACTION	UNLOOP
	$classifi:Classification	$latestlabresult:LatestLabResult	$otherlabresult:LatestLabResult	$otherclass:Classification	$p:Patient	$drs: DiabeticRiskScore		
	problemCode == "$param"	patientId == $classifi.patientId, labCode=="$1", labValue $2	patientId == $classifi.patientId, labCode =="$1", labValue $2	patientId == $classifi.patientId, problemCode == "$param"	ethinrange(ethCode), patientId==$param	patientId==$param,riskScore <4	$drs.riskScore = $param; update($drs);	
Recall Decision	**Check problemCode**						**decide recall**	**no-loop**
DM_COD	**DM_COD**	**44TB,>55**				$classifi.patientId	**$drs.riskScore+1**	**TRUE**
DM_COD		**2469,>130**	**246A,>80**			$classifi.patientId		**TRUE**
DM_COD		**46TD,>3**				$classifi.patientId		**TRUE**
DM_COD		**44J3,<60**				$classifi.patientId		**TRUE**
DM_COD		**44Q,>=1.7**	**44P, >=4.0**			$classifi.patientId		**TRUE**
					$classifi.patientId	$classifi.patientId		**TRUE**
DM_COD				CSMOK_COD		$classifi.patientId		**TRUE**
	DM_COD			heartdisease		$classifi.patientId	**4**	**TRUE**
				stroke		$classifi.patientId		**TRUE**
				TIA		$classifi.patientId		**TRUE**
				PAD		$classifi.patientId		**TRUE**

Figure 4.9: Diabetes risk score of a patient.

Best Practice Advocacy Centre New Zealand, a factor should be considered as a risk factor only if it directly applies to the current health condition of the patient. From a

Table 4.5: Diabetic care action plan.

Review Name	**Low Risk**	**Medium**	**High Risk**	**Very High Risk**
Clinical Review	6	3	3	3
HbA1c	6	3	3	3
Blood pressure	6	3	3	3
Lipids	12	12	12	12
ACR	12	6	6	6
eGFR	12	3	3	3
Foot check	12	12	6	3

workload point of view, this diabetic risk score of a patient determines the frequency of recalls for the patient, where low risk patients are on a 6 month recall, and medium, high risk and very high risk patients are on a three month recall (refer to the first row in Table 4.5).

4.5.4 Decision table: to decide recall period for a patient

From the discussions above, we can conclude that based on the severity of a specific LTC in a patient, we could decide the frequency of recalls required to manage that LTC in a patient. But, for multi-morbid patients more than one care plan applies. Also, from the dataset it is difficult to have an explicit understanding of the care plans applied to the patients. For example, for a given patient A suffering from multiple LTCs, say, condition_1, condition_2, and condition_3, currently it is difficult to draw a conclusion, from the data, on how the recall decision is made. This makes it difficult to develop a rule that directly decides the recall frequency required for a patient to provide the best practice care. Therefore, deciding the recall period for a patient (to be used in simulation) involves a two-step process as follows.

Step 1 Find, for each patient:

(i) which LTCs are present. As presented in Section 4.5.1, we map the classification Read codes to generic names. This help us identify which LTCs are present in them;

(ii) identify the corresponding recall periods for the LTCs present in them. We implemented care plans for three selected LTCs, namely diabetes, CKD and hypertension. As mentioned in Section 3.2.2, the scope of this work is limited to producing a proof-of-concept workload model. Therefore, we applied a 6-monthly recall period for all the other LTCs in Section 4.5.1 as well as to those patients who are not associated with a care plan for the listed LTCs but have made an LTC appointment historically. Thus, we assign a recall period for each LTC in a patient based on the care plans that apply to them.

Step 2 Decide from the multiple recall periods of a patient, which recall period drives their recalls. As the severity of a condition worsens, frequent interventions are required to manage that LTC in a patient (Best Practice Advocacy Centre New Zealand, 2012). Hence, we assume that the most frequent period among multiple recalls periods for a patient drives their recalls and we therefore calculate the recall period as simply the minimum of the recall periods of a patient's LTCs.

Table 4.6: Individual Recall table

idno	careplan_code	recallinfreq
A	DM_COD_PLAN	3
A	heartdisease_PLAN	6
A	CKD_PLAN	12

For instance, refer to Table 4.6. In our context, the patient A would be recalled every 3 months. Later in Chapter 5, I will explain how these individual recall decisions are encoded as rules in our workload prediction model to determine a patient's recall period and simulate the workload.

Having presented the pre-processing steps, the following section describes the initialisation of our workload prediction model.

4.6 The initial patient model for simulation

Preprocessing of the data involved "cleaning" the data to fill the gaps and adopting assumptions and definitions that help us to infer information required for this study. In this section, I explain the preparation of data as an initial patient model for our simulation. Our first step is to identify the cohort of patients. Once the cohort is

decided, the associated LTC appointment details of these patients are required to drive the simulation process to simulate patient visits. The PostgreSQL queries for the preprocessing are appended in Appendix A.10. Please remember that these queries will include MedTech32 specific tags and hence will need to be updated according to one's dataset, if trying to reproduce the results following the steps discussed in this thesis.

4.6.1 Our cohort

Following the preprocessing step, an LTC patient associated with a recall period is considered for simulation. However, based on our simulation period, we further filter these patients to those who have made an appointment historically for the simulation period (this step helps us to compare the historic workload to the predicted workload from the same set of patients and for the same period of time).

4.6.2 The LTC appointments that drive the simulation

Following the steps described in Section 4.4.1, we identified the LTC visits of the cohort patients. The next step is to identify which LTC appointment should be considered as a starting point to decide the recalls for the patient during the simulation period. Given there are multiple LTC appointments for a patient, there are two ways to handle this.

Method 1 Consider the most recent visit that falls prior to the simulation start date. However, we found out that vast majority of the patients (irrespective of their recall period) have their last visit for the year in the last quarter of the year (refer to Figure 4.10). This could be mainly due to the forthcoming (southern hemisphere) summer holidays. In the figure, we have not included patients with 1 and 2 month recall periods, as they would have a visit in the last quarter anyway. The 3-monthly recalled patients will have a visit in the year end; for 6 monthly patients the Q2 and Q4 patients will have visits in Q4; and, all 12 monthly patients in Q4 will have visits in Q4 in our simulation. Hence, choosing the most recent visit prior to the simulation start date will have a bias that would accumulate simulated visits towards the end of the year, as observed (see Figure 4.11) in our initial results with cohort of patients with CKD only.

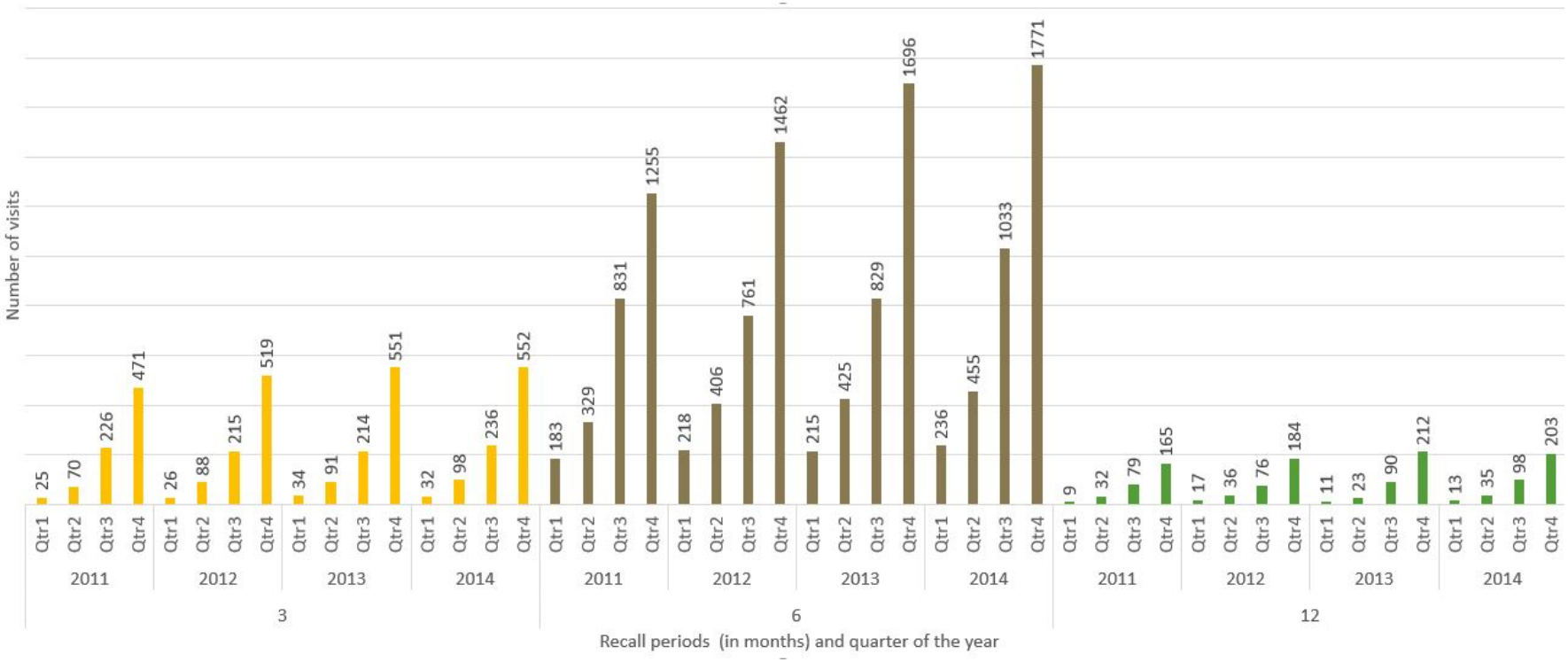


Figure 4.10: Trend of last visits of patients per year categorised by their recall period and quarterly.

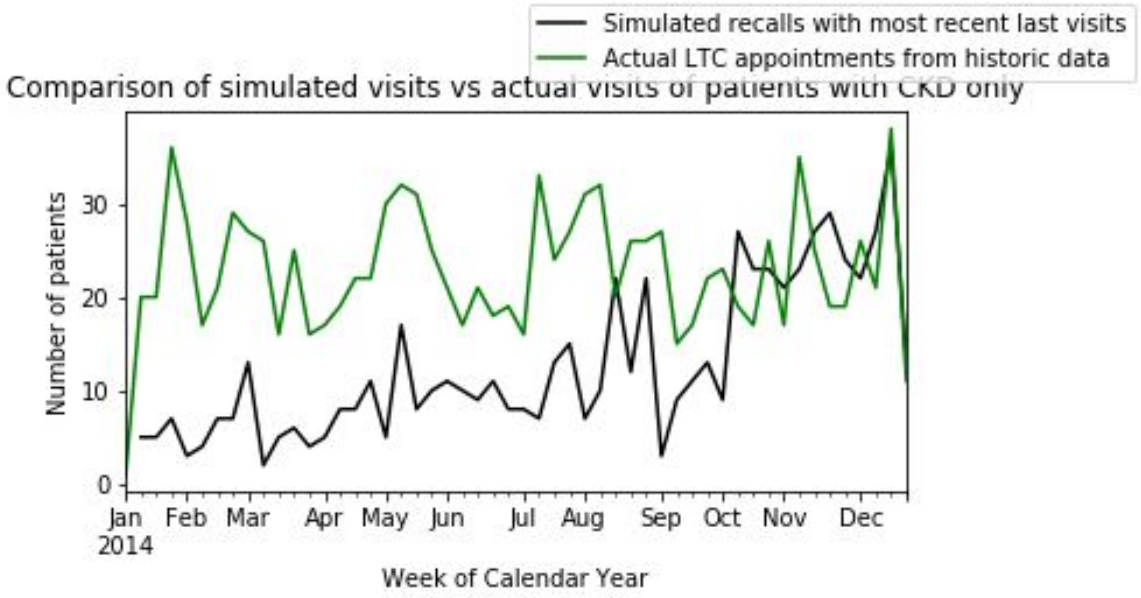


Figure 4.11: Comparison of simulated visits of patients with only CKD (using their most recent visit in 2013) vs actual LTC visits for the year 2014.

We also observed that there were more frequent visits from patients who would visit less frequently as per the care plans (refer to Figure 4.12 see 6-monthly and 12 monthly recalled patients number of visits). And, some patients visited less frequently than required by their care plans (refer to Figure 4.12 see 1, 2 and 3 monthly recalled patients number of visits). There were patients who made frequent visits during the first half of the year, and did not visit the practice for the rest of year. Therefore, we chose a generalised Method 2 described below to choose the visit of a patient that would simulate their subsequent recalls.

Method 2 When there are multiple LTC visits in the year prior to our simulation start date, we randomly chose one visit as their last visit of the year that drives their timely recalls.

It is evident from Figure 4.13 that randomly choosing the last visit would break that periodic pattern of visits that accumulates towards the end of the year (for 2013).

Hence, our initial patient model has LTC patients with their LTC recall periods and an LTC appointment date which is considered as the starting point for timely recalls for them in our simulation. We also did a more sophisticated approach using

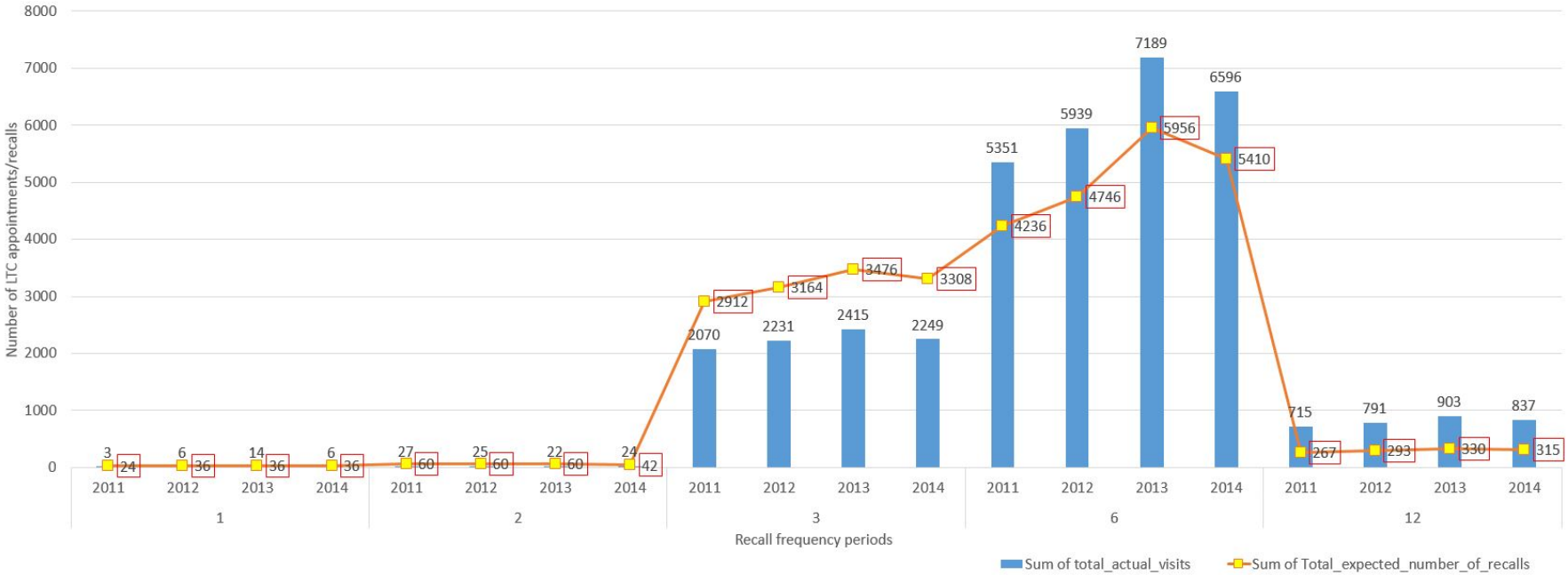


Figure 4.12: Comparison of number of expected visits as per care plan and actual number of visits categorised by year and morbidity.

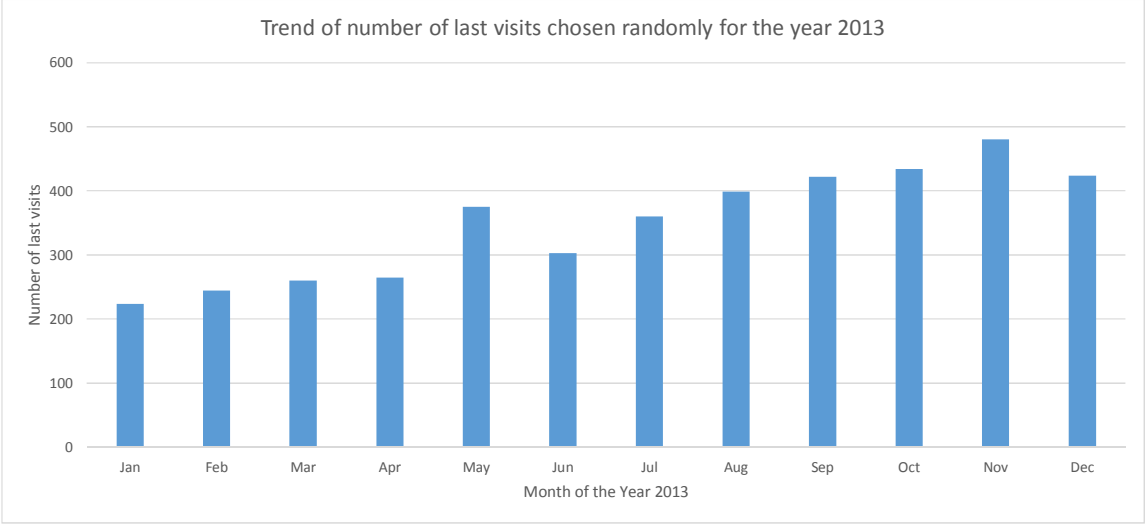


Figure 4.13: Trend of randomly chosen last visits of patients for the year 2013.

Bayesian inference to decide the visit pattern of patients (see Chapter 7 for more detailed discussed on using Bayesian inference in our workload prediction model.)

The following section will describe our approach to validating a few assumptions, eGFR calculation and rules of our rule set.

4.7 Validation of assumptions, eGFR calculation and the rules

Validation of assumptions

One of the assumptions that we used in our preprocessing step was that the LTC medication prescriptions may be valid for four months (although they are prescribed three-monthly). We validated this assumption by checking, for each patient, that these LTC medications (which are T-tagged, see Section 4.4.1) are prescribed at least three times in the previous year from the (considered) date of LTC appointment of the patient. We also manually checked that those medicines that were not prescribed at least three times in the past were prescribed in the last six months from the date of appointment, which accounts for either a new diagnosis of an LTC, a change of medication for an LTC, or a change of severity of an LTC in the patient. From the results, we found that it is a reasonable assumption that the LTC medications are prescribed at least three times in the previous year of the date of appointment. However, we also observed

that there are certain medications that are usually prescribed for acute conditions that were included in the T-tagged medications list. According to our PCA, there are cases where medications for acute cases such as paracetamol may be prescribed as part of a bulk prescription for LTC medications. Therefore, we chose to include all T-tagged medications, although some of them may be for acute problems in the patient.

Validation of the eGFR of a patient

According to the National Kidney Foundation (2015), "*GFR is the best test to measure your level of kidney function and determine your stage of kidney disease. Your doctor calculates it from results of your blood creatinine test, your age, body size, and gender.*" Hence, apart from the serum creatinine and ACR tests, doctors may advise GFR (Read coded using "451F") tests (National Kidney Foundation, 2015). However, not all GPs request GFR (451F) lab tests, as the standard way of measuring kidney disease used to be through serum creatinine (44J3) tests. According to the labs data for our cohort, there are 2 531 patients who had lab test results recorded for their GFR (Read code: 451F) and serum creatinine (Read code: 44J3) values. Out of these 2 531 patients, only 931 patients are on CKD plan, which seemed to be a significantly low number to our PCA. Therefore, our CKD care plan involved (for each patient) calculations of the Gfactor (from the 44J3 lab results) and ACR stage, and then using them together to decide the CKD stage, which turns out that our cohort has 1 498 CKD patients.

In order to validate our eGFR calculation, we cross-checked the readings for 451F on a date for a patient with the corresponding eGFR value calculated using the function in Listing 4.3 for the (931) patients, for whom the results were available. We found that except for a few cases (nine patients in their advanced CKD stage), the values matched. Patients in their advanced CKD stage (G5) decrease their eGFR dramatically within a year (National Kidney Foundation, 2015). However, as we are not simulating patient outcomes and, also simulating the steady state of the practice, we chose to include them in our cohort.

Validation of preprocessing rules

Identifying LTC patients involved mapping Read codes to generic problem classes. Hence, a measure adopted to validate the decision table that maps Read codes to the respective problem class(es) was to check whether the number of patients with LTCs was identified correctly. In order to check the number of patients with LTCs, we generated the multi-morbidity matrix given in Table 4.7. As we do not have access

Table 4.7: Multi-morbidity matrix.

Care plan code	asthma	Atrial	CKD	COPD	Diabetes	heartdisease	heartfailure	Hypertension	PAD	psychosis	stroke
asthma	1406	52	177	45	92	83	20	174	5	8	18
Atrial	52	393	253	43	68	148	73	114	20	2	42
CKD	177	253	1498	130	342	406	131	0	62	8	102
COPD	45	43	130	288	50	81	20	78	19	1	18
Diabetes	92	68	342	50	640	159	45	0	32	7	42
heartdisease	83	148	406	81	159	764	93	281	43	4	60
heartfailure	20	73	131	20	45	93	162	22	15	2	15
Hypertension	174	114	0	78	0	281	22	1399	24	9	67
PAD	5	20	62	19	32	43	15	24	97	0	15
psychosis	8	2	8	1	7	4	2	9	0	46	1
stroke	18	42	102	18	42	60	15	67	15	1	190

to the actual database, we shared this multi-morbidity table with our PCA[16], who confirmed that it is a "reasonable" result and that our results appeared suitable as input to continue with our workload prediction phase. In summary, due to the database available not including certain information that was needed, we had to preprocess the anonymised dataset and further prepare the data to be seeded into our workload simulation model. After these data filtering and preparation steps, the final dataset i.e., our initial patient model, included:

- the cohort patient details with their recall periods per LTC; and,
- the LTC appointment date of each patient that would be considered as the starting point to drive their recalls (see Subsection 4.6.2).

In the following chapter, I will explain how our initial patient model drives the simulation of patient visits, and how the workload is then aggregated to a weekly basis for a year. That chapter will also revisit some of the assumptions and definitions presented in this chapter.

[16]Only the PCA has access to the original database.

Chapter 5

Workload Prediction: The Simulation Process

In the previous chapter I presented the initialisation phase which gives the individual recall details, and explained how the associated LTC appointments are identified for the given patients. Given the patient visits and the recall frequencies corresponding to their LTCs, our adaptable best practice based workload prediction model (ABP-WPM) simulates patient visits. These patient visits are aggregated at a population level on a weekly basis to give the workload due to LTC patient visits at the medical practice for a specified future period (set to one year in this thesis). We will first discuss our assumptions in the context of a patient visit and then discuss the workload prediction process. Next, the results of workload simulation are presented, where we also discuss a few unanticipated challenges in predicting workload for the medical practice following the best practice guidelines. Furthermore, we also discuss a few variations in the guideline rules to adopt the best practice for this cohort of patients. For example, we assume a six-monthly recall period as a baseline recall period, which means an LTC patient will be seen at least once in six months.

Assumptions

There are a few assumptions we make in order to design the rules for our simulation. These assumptions are listed below.

- Multi-morbid patients need to be seen more often to manage their LTC management needs. Hence, we assume that their most frequent recall period should be considered to decide the follow-up appointment;

- We assume the patient population at the practice is in a steady state. Thus, during simulation we do not attempt to predict changes in patient conditions, or patients joining or leaving the practice. In other words, during simulation each patient represents a prototypical patient with certain conditions and not a specific patient at the practice.
- We assume that patients turn up for their recalls as required by their respective plan-of-care, and that they turn-up exactly on their recall date (with those recalls that fall on a holiday pushed to the next working day, but we explore the effects of relaxing this assumption in our evaluation).
- We assume that the practice attends to LTC patients only on weekdays.

A few variations on these assumptions and their impacts are discussed in Chapter 6 with a few what-if scenarios.

5.1 Workload prediction process

Let us consider an example of three patients, A, B and C in our initial patient model from our preprocessing step, with their last visit dates in the year prior to simulation start date and their most frequent recall periods calculated. The simulation period is the year 2014. The following steps are summarised in Figure 5.1 which depicts three patients with their chosen last visit date and most frequent recall period[1]. Their next recall date is calculated as:

$$recallDate = LastVisitDate + MostFrequentRecallPeriod$$

Then each of these recall dates are compared to our simulation start date. If the recall date falls prior to our simulation start date, then the most frequent recall period is repeatedly added to the last visit date until the recall date falls on or after our simulation start date (shown in the recalls_initialisation block in Figure 5.1). These updated recalls then initiate patient visits within the simulation period. We assume that when a recall is scheduled, patient visits occur on that recalled date. When a patient visit occurs, the subsequent recall date is scheduled as follows:

$$recallDate = VisitDate + MostFrequentRecallPeriod$$

[1]the smallest recall period, e.g., 3 months recall is smaller, and hence yields more frequent recalls, than six months.

This is shown in the recalls_simulation block in Figure 5.1. This scheduling of recalls and corresponding patient visits happens until the simulation end date. When the simulation ends, the recall dates in the model are mapped to a calendar year. This is highlighted with the Patient Ids (A, B and C) in the calendar in the figure. As seen in the figure, the visits that fall on a weekend or a holiday are pushed (by a rule) to the next working day. For example, the visit of A that falls on Sunday the 13^{th} of July is pushed to the 14^{th} of July. Then on the 14^{th} of July, patient A visits, and a new recall date is calculated. This calculated recall happens to fall on the 14^{th} of September which is, again, a Sunday and hence is pushed to the 15^{th} of September. Then, the recalls are aggregated as given in the aggregated workload graph in the figure. Although, in this figure, the aggregation is shown on a monthly basis, in this work we aggregate the workload on a weekly basis.

In short, with the individual recall details and LTC appointment dates, the rules in our model instantiate a care pathway in an iterative fashion, for each patient, visit-by-visit, resulting in a plan-of-care for the patient. This gives an anticipated frequency of required GP appointments for each patient over a period (a year, in this study). Aggregating this, for all the patients, gives the overall workload for the specified time frame, on a weekly basis. The following section presents the rules that simulate the patient visits and calculate the corresponding recalls. The rules are presented in Drools *.drl* format.

5.2 The rules and the workload simulation

In order to initiate the simulation, we have Today (the current day during the simulation period), IndividualRecallDetail (that records the recall period for a patient), PatientVisit (which is an event that simulates the consultation with a role in the practice) and Recall (records the recall date of a patient, and the role for which the recall is scheduled for a patient) facts in the working memory. Please note PatientVisits are equivalent to Recalls, in a way that recording either one of them would give the predicted workload. We chose to record Recalls, and delete the PatientVisit once a corresponding Recall is created by the PatientVisit.

Our simulation starts on a *SimulationStartDate* and runs until *SimulationEndDate*. On each day of simulation, if the conditions of the rules hold, the rules listed below are fired (refer to Algorithm 1). Please refer to Appendix A.9 for their Drools implementation.

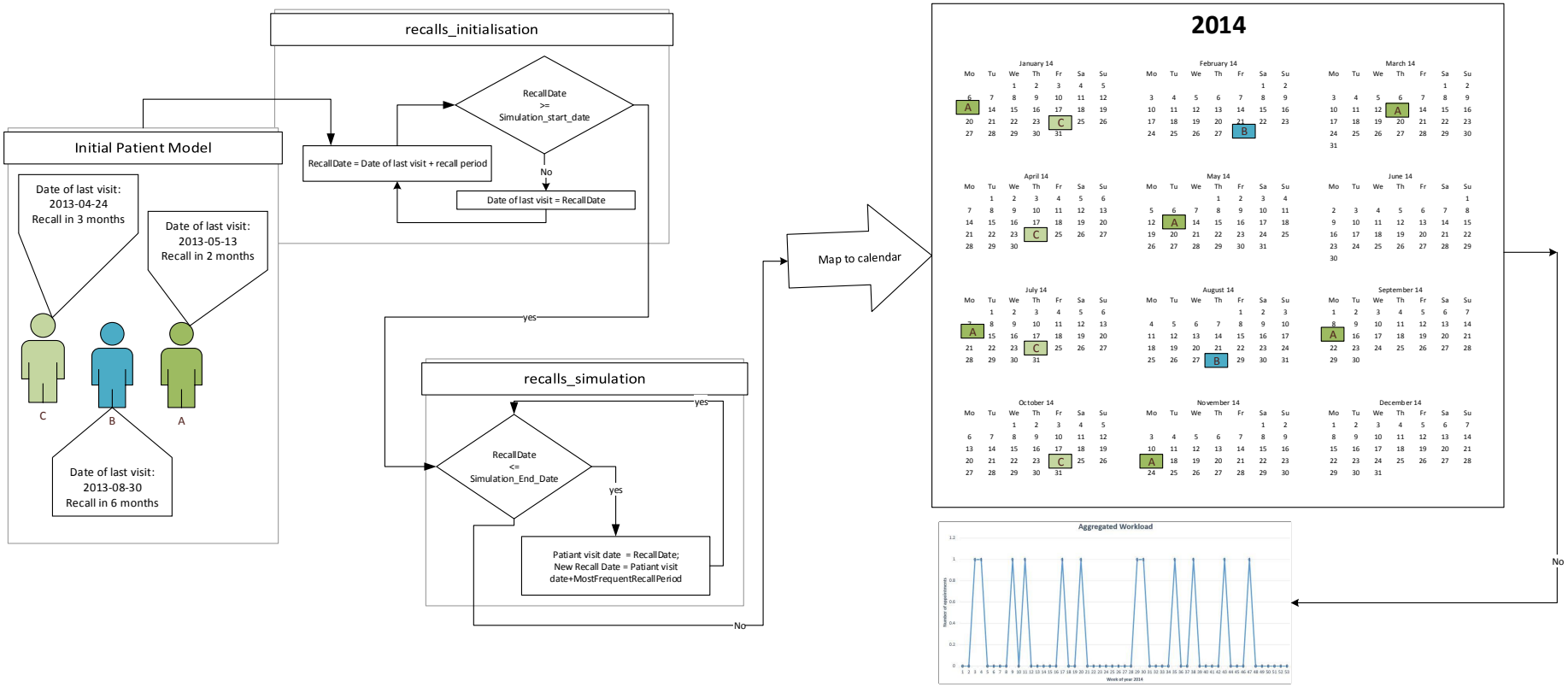


Figure 5.1: Overall process showing inputs and how the recalls are mapped to a calendar, and then aggregated to weekly workload.

The following lists the rules used in our simulation:

- Rule to create a patient visit when the recall date is reached for a patient. On a day in simulation, if there are recalls scheduled to be on that day, then patient visits occur (see Section 5.2.1).

- Rule to create a recall date when a patient visits. In a real-world scenario, we expect that the date for a follow-up for LTC patients will be discussed and agreed on during the patient visit. So, in our rules we also create a next recall date on a patient visit (see Section 5.2.2).

- Rule to update the recall date to fall on a working day. We assume that LTC patients are consulted only on business days of the practice (see Section 5.2.3).

In the algorithm we have presented the conditions in the order they would be executed in the rule engine. In the actual rules discussed below, we use the *salience* attribute of the rule to set the priority for each rule. Then, higher the salience value, the higher the priority. Having these rules layered with priority eases the capability to extend the rule base for our what-if scenarios presented in Chapter 6.

```
Result: today = SimulationEndDate + 1
today= SimulationStartDate;
while today <= SimulationEndDate do
    if recall date for a patient==today then
        create Patient Visit for that patient with today as the date of visit;
    end
    if date of visit for a patient==today then
        Get the least frequent recall period (in months) of the patient;
        next appointment date = today + the least frequent recall period;
    end
    if the next appointment date for a patient is a holiday then
        next appointment date = next appointment date + 1;
    end
    today = today + 1
end
```

Algorithm 1: The encounter based unfolding of a plan-of-care for an individual on a simulation day within our adaptable best practice based workload prediction model.

The Drools Fusion CEP[2] rule engine has an inbuilt "*pseudo-clock*" that can be accessed and advanced through its in-built function to add hours, minutes or seconds or days, months or years too. Our simulation walks through each day (of the year 2014). During the patient visit on their scheduled recall i.e., the encounter based unfolding of a plan-of-care for an individual, the next recall is scheduled. Algorithm 1 presents this walk through of the events on a simulation day. "today" refers to the current day of simulation within the simulation period. If there are recalls scheduled for "today", the corresponding patient visit occurs, and thus, the patient's date of visit is "today". A corresponding recall date is created from that date of visit. This continues until no more rules are activated. Then the clock advances to next day.

5.2.1 Rule to create a patient visit when the recall date is reached for a patient

When the recall date of a patient is reached within the simulation period, a patient visit occurs (we denote it by a PatientVisit fact in the working memory, see line 7 in Listing 5.1). This fact depicts patient visits (the event of a visit of a patient). Having it as a separate fact can give more flexibility for our model. For example, this rule enables us to record if a recall is scheduled for a particular role (e.g., a GP) and later if the patient is required to visit another role outside the practice such a physiotherapist. In this work we focus on recalls with a GP or a nurse (roles within a practice), so recalls and patient visits are complementary.

Listing 5.1: Rule to create patient visit when the recall date is reached for a patient.

```
1 rule "If working day today, and recalls are there, then create
    patient visits."
2 salience 5000
3 when
4   $t:Today()
5   $recalltoday:Recall(recallDate.getTime() == $t.today)
6 then
7   PatientVisit $newpv = new PatientVisit();
8   $newpv.patientId = $recalltoday.patientId;
9   $newpv.dateOfVisit = $recalltoday.recallDate;
10  insert($newpv);
```

[2]complex event processing

```
11 end
```

5.2.2 Rule to create a recall when a patient visit happens.

In a real-world scenario, we assume that a GP considers all the information about the patient to decide the next follow-up. Hence, during a patient visit, our simulation rules collect the recall frequencies for with different care plans associated with that patient, and use the most frequent (minimum) recall period, if there is more than one recall period for that patient. I use the Drools CEP's *accumulate* feature to gather and take the minimum value among all the *individual recall details* of the patient (*patientId*) to decide the next follow-up date for that patient as given in Listing 5.2.

This rule reads as follows: when it is *Today*, and a patient visits whose individual recall details are available, and their minimum recall frequency ($minrecallInFreq, see line 5 in Listing 5.2) can be calculated by accumulating all the individual recall details for this patient (see line 4 in the listing), then create a new recall for the patient with a new recall date as *date of visit* ($t) plus $minrecallInFreq (line 11 in the listing). Insert this new recall into the working memory (line 12 in the listing) and remove the patient visit from the working memory. We need to record either patient visit dates or the recall dates for later aggregation. I chose to keep a record of recall dates. Hence, once the next recall date is scheduled, the patient visit is retracted from the working memory (line 13 in the listing). Please note that this rule is later extended to choose a deviation from the expected recall date (see our Bayesian approach in Chapter 7).

Listing 5.2: Rule to create patient recall when a patient visits

```
1  when
2   $t:Today()
3   $pv:PatientVisit(dateOfVisit.getTime() == $t.today)
4   exists (IndividualRecallDetail(patientId == $pv.patientId))
5   $minrecallInFreq:  Number (intValue > 0) from accumulate ($ind :
      IndividualRecallDetail(patientId ==
      $pv.patientId,$recallinfreq:recallInFreq), min($recallinfreq)))
6  then
7   long min_freq = (long)$minrecallInFreq.intValue();
8   Recall $newrecall = new Recall();
9   $newrecall.patientId = $pv.patientId;
10  $newrecall.recallInFreq = min_freq;
```

```
11  $newrecall.recallDate =
      java.sql.Date.valueOf($pv.dateOfVisit.toLocalDate().plusMonths(min_freq));
12  insert($newrecall);
13  retract($pv);
14 end
```

5.2.3 Rule to update a recall date that falls on a holiday day

Once the recall date for a patient is created, the rule evaluates whether this recall date is a working day or not. If it is not a working day then, the recall date is updated to be in the next day in the calendar year. Please remember that the rules are first evaluated against the facts and those matched rules are added to a list of activated rules. Hence, this rule's conditions are written in such a way that it checks that all the recall dates that fall in future are on a working day. All the recalls that fall on a holiday are pushed to the next day (the line 11 in Listing 5.3).

Listing 5.3: Rule to push recall date to a working day

```
1  rule "Push recalls to next week if recalls falls on a holiday"
2  salience 5000
3  when
4    $t:Today()
5    $recalltoday:Recall(recallDate.getTime() > $t.today)
6    not(eval (workingday($recalltoday.recallDate.getTime())))
7  then
8    Recall $newrecall = new Recall();
9    $newrecall.patientId = $recalltoday.patientId;
10   $newrecall.recallInFreq = $recalltoday.recallInFreq;
11   $newrecall.recallDate =
      java.sql.Date.valueOf($recalltoday.recallDate.toLocalDate().plusDays(1));
12   retract($recalltoday);
13   insert($newrecall);
14 end
```

The function *workingday()* takes a day and checks if it belongs to a list of holidays (New Zealand public holidays in this case), or if it is a weekend. In either case, workingday() returns False, otherwise it returns True. For every recall for which the eval is False (line 6), the recall is pushed to next day (line 11 in the Listing) which re-evaluates the rule with the new recall fact. This runs multiple times so if the recall

is Sunday, and adding one day returns Monday, and Monday is a public holiday, then it rolls over to Tuesday.

5.3 Results and Discussion

We know from the literature that adherence to clinical guidelines is suboptimal at the primary care level (Dee Mangin, 2012; Fischer et al., 2016; Haynes and Haines, 1998). Therefore, there will be a difference between the predicted workload following the best practice and the workload attended historically (we refer to this as the capacity mismatch). Our dataset, too, is not different in that perspective. We compared the number of actual LTC visits (according to the LTC definition given in Section 4.4.1) and the number of visits required if best practice for LTCs is followed for these patients. Figure 5.2 shows that patients who must be seen more frequently (1-monthly, 2-monthly and 3-monthly recalled patients) visited historically less frequently than required by their care plans, and patients who need to be seen less frequently visited more often than required by the care plans applicable to them (see the six-monthly and 12-monthly recalled patients in the figure). This implies that there will be a difference when the predicted workload is compared to the historical LTC appointments of the patients.

More complex the LTC needs are in a patient, more frequent the visits must be. We also found it interesting to see (refer to Figure 5.3) that the average number of visits per patient per year is the same irrespective of number of LTCs in a patient (except for the year 2012 for patients with 11 LTCs in them). This highlights that the historic recall data does not adhere to the best practice guidelines, as some patients who are expected to visit less frequently visit more often, and on the other hand, patients who are expected to visit more frequently, visit less frequently than expected. These historic visits may due to requirements other than their LTC management; for example an acute exacerbation of an LTC in a patient.

The best practice at a PHC not only aims to adopt the guidelines, but also adapt according to what works best for its patients (Johnson, 1997). Hence, at the particular healthcare centre whose data is used in this study, according to its Clinical Manager, every LTC patient is seen at least once in six months. Therefore, we updated the 12-monthly recalls for LTCs to a six-monthly recall (please bear in mind that this may vary across practices). This variation in the guidelines made our model align better to the actual workload that occurs at the practice. We emphasise that the aim of this workload prediction model is to predict the upcoming workload from the patients who

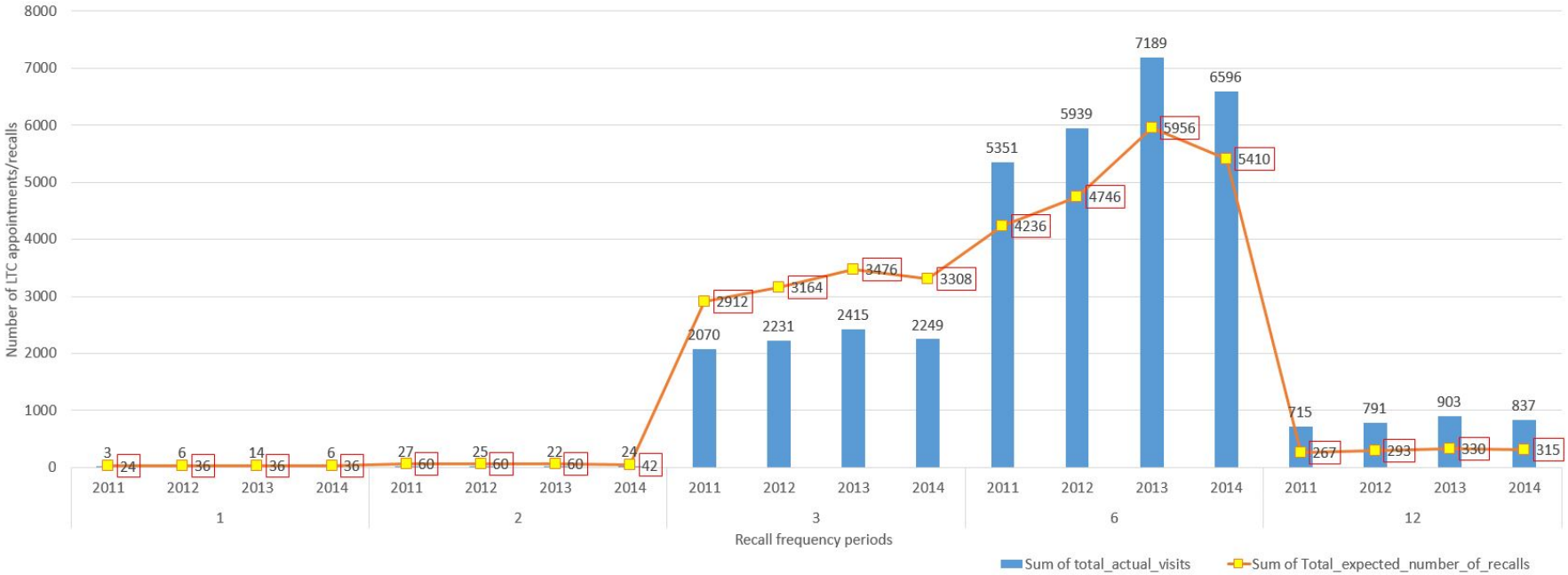


Figure 5.2: Comparison of (per year) total number of actual LTC visits and expected number of LTC visits from LTC patients (categorised by their most frequent recall period) whose workload is generated for 2014.

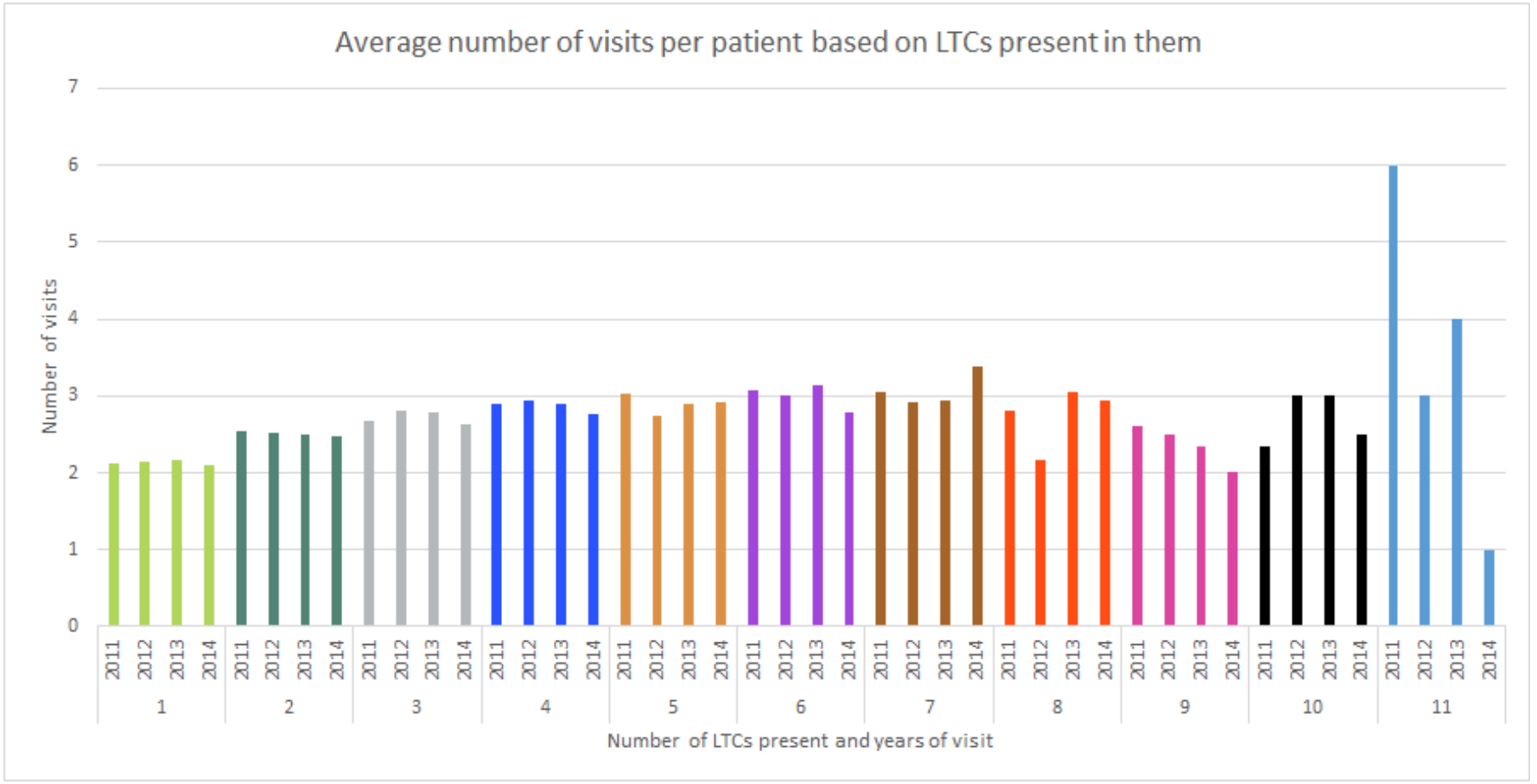


Figure 5.3: Average number of visits per patient per year differentiated by number of LTCs (multi-morbidity).

are recalled according to their care plan requirements, and not to match the predicted workload to the actual workload that was observed at the practice in the past, and, at the same time not to rigidly follow guidelines without considering the reality.

The workload simulation model was built through DSRM cycles discussed in Chapter 3. In this section, I discuss the results of simulating patient visits of all the LTC patients (corresponding to Cycle 4 of Figure 3.5).

For this phase, our initial patient model from the preprocessing step included:

1. patients who had face-to-face consultations in both 2013 and 2014. In our dataset, we have patient details where they are prescribed medications over phone. Our PCA was interested in the workload which accounts for GP time slots. We also needed a measure to compare our predicted workload and its impact on the practice. In order to do this comparison, we chose only those patients who had visited both in 2013 and 2014. Therefore, although, 6 154 patients had visited in 2013 and in 2014, including visits to collect medication prescriptions, we consider only the 4 190 patients who have made face-to-face consultations during both 2013 and 2014.

2. for each patient, a chosen last visit date to seed the simulation (see Section 4.6.2).

3. all individual LTC recall periods (with all 12-monthly individual recall periods for LTCs updated to six-monthly periods).

Within the simulation we simulated recalls that fell prior to the simulation start date. For instance, a patient whose chosen visit to initiate their recall may be in June 2013, and their corresponding simulated recall may fall in December, 2013. We tried two ways to handle such recalls that fall prior to simulation start date.

Method 1 Ignore those patient visits that fall prior to the simulation start date. It is evident from Figure 5.4 (the bottom line in the figure, the *Simulated recalls - updated assumptions*) that the simulated workload in this case is much less than the actual workload.

Method 2 Assume that they followed their best practice during their visits in 2013, and roll over their visits to fall within the simulation period. Then the predicted workload aligns with the actual workload of the practice. There are two reasons why this is required. Firstly, as discussed in Section 4.6.2 we are randomly choosing one visit among multiple visit, it is therefore possible that the chosen visit was the

one from the first half of the year (so it need to be extrapolated to fall within the Simulation period). Secondly, it is possible that patients would have visited less frequently than required and those visits could have been in the first half of the year. (In Figure 5.4 the *simulated recalls corrected* corresponds to the workload including this variation in patients' recall periods).

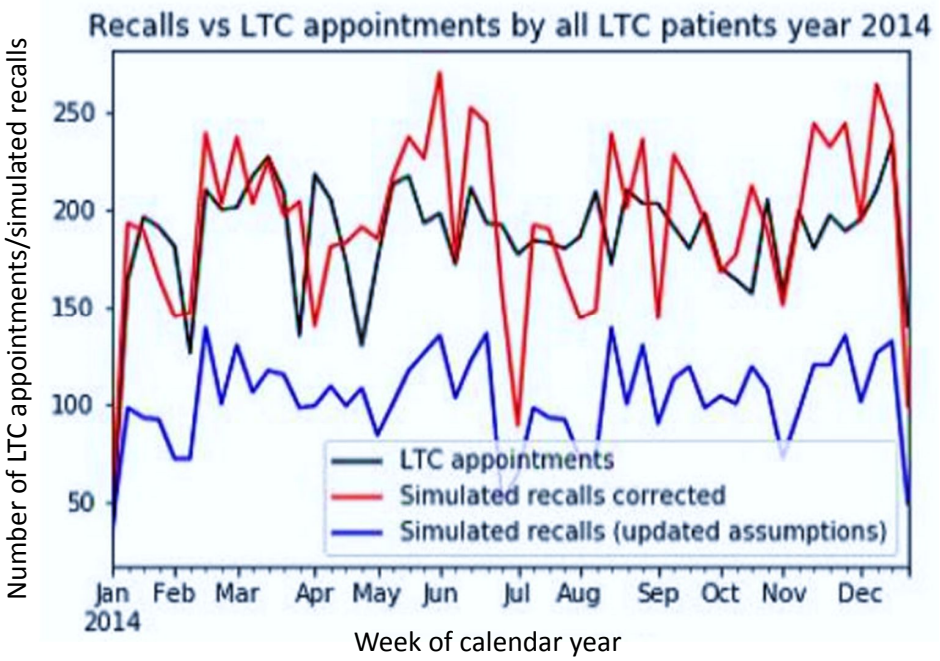


Figure 5.4: Comparison of simulated visits, the 12-monthly recalls updated to six-monthly recalls and choosing a last visit date from multiple visits for each patient over the last year, vs actual LTC visits of cohort patients for the year 2014.

I emphasise again that the aim of this work is not to make the prediction align with what happened but to give a sense of what would happen when the practice follows best practice guidelines for LTC patients. Thus, the simulation model takes into consideration both the evidence of visit (i.e., considering only those patients who have visited during both the years, 2013 and 2014) and the best practice guidelines to decide recalls to anticipate future workload.

5.3.1 Patient visits of patients registered with a particular GP

One measure we used to evaluate our model, in order to try and reconcile the difference between not following best practice guidelines and historical approach for care, is by selecting a cohort of patients of particular GP who, we were adviced by the Clinical

Manager of the practice, did follow best practice for his LTC patients. Multi-morbidity (a combination of these LTCs) would influence the visit pattern of LTC patients and is complex. Therefore, since we have implemented only three care pathways (i.e., for CKD, diabetes and hypertension) and need to avoid patients with multi-morbidity, we chose to look at patients who have only one of these three LTCs and are registered with this GP. This constraint helps us to be more accurate about their visit patterns. Although the historic number of visits per patient matches (in the sense the total number of actual visits in a year matches with the number of visits predicted through simulation given in Figure 5.5), the cohort (three patients) is not big enough to draw a conclusion.

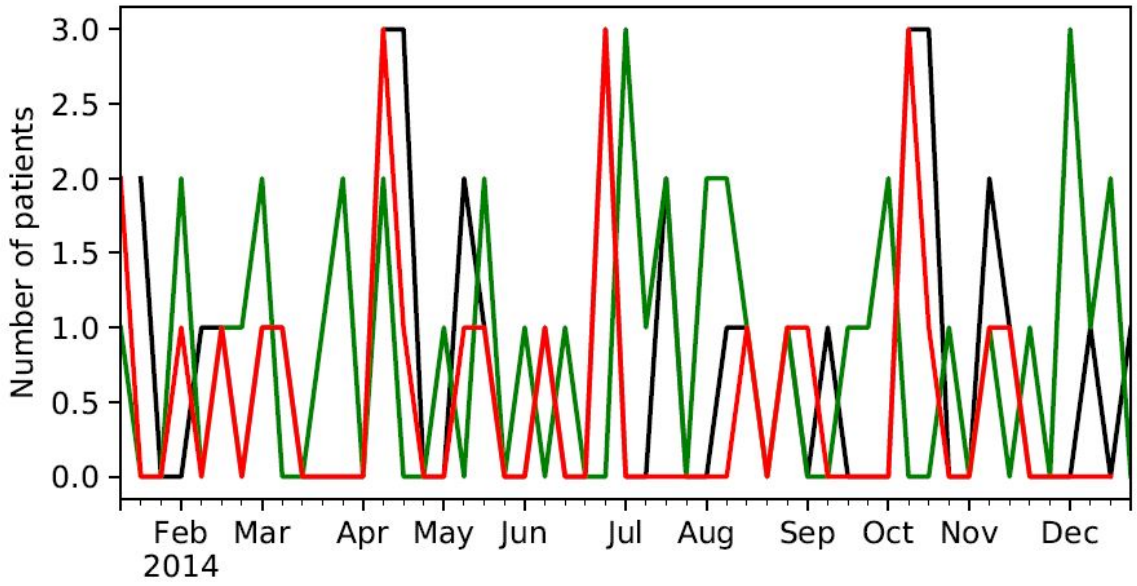


Figure 5.5: Patient visit vs simulated visit for patients registered with a specific GP.

Chapter 6

What-if Scenario Analysis

In the previous chapter, we saw how care pathways unfold into a plan-of-care to follow the best practice guidelines to manage LTCs in a patient, and we focussed, in particular, on their frequency of recall. In this way we were able to predict the workload at a PHC centre as a volume of number of visits from its LTC patients. Until now, we were considering a PHC as a whole and the workload was addressed to the GPs at the practice. However, from a practitioner's[1] point of view, they need a measure to manage their workload within the practice's resources, which also include nurses in the care planning and delivery process.

In this chapter, therefore, we will consider various roles, specifically the GPs and the nurses who can attend their LTC patients. We use using what-if scenarios to explore how knowing the workload can help a PHC manage the LTC workload. This way knowing the population-level workload can help make a shift from patient-initiated reactive care to a proactive LTC care, considering various alternatives of practices for managing LTC patients.

We also explore how knowing what the health strategies have in-store for a nation-wide population could help a PHC centre understand and handle the impacts of such health schemes. This foreseeing of the impacts and taking measures to handle the changes is addressed using various names such as process (re)engineering, and change management. In this thesis, we call it "what-if scenario analysis".

Broadly, the factors that impact workload management could be within the organisation, such as availability of various roles, external factors beyond the control of organisation such as a flu outbreak that adversely impacts the workload, or the im-

[1]The PCA, whose feedback drives our assumptions and definitions, is a GP himself. In this work, we consider him as a representative of the GPs' viewpoints.

pact of new policies or procedures. According to the Planning and Funding Clinical Director of Primary care, Waitemata, a change in policy or procedure can impact an organisation in two ways[2]:

1. Clinical Risk Analysis: Whenever a policy changes, the patients and their care providers are affected equally. The GPs have a better understanding of the impact on the care delivered than the volume of patients affected by the change in policy. Therefore, usually GPs do not consider the impacts of policy changes at the practice.

2. Financial Risk Analysis: This considers the financial aspect of the organisation. The New Zealand Ministry of Health funding schemes are complex, involving direct funding, funding through public health organisations (PHOs) and the district health boards (DHBs), and are designed around specific health requirements and the health care level that address these requirements (see Figure 6.1). These are reflected in the nation's health strategies too. For example, a Ministry of Health NZ (2016a) scheme provides funding to address diabetes and prevent associated complexities. However, a PHC organisation lacks a holistic view of how this funding scheme would impact meeting other patients' LTC management needs, and what are the financial aspects the organisation should consider to implement these health strategies to gain their full potential. Based on this feedback, therefore, although we do not consider this to be a complete financial model, we attempt to address the impact of change on the practice in terms of the co-payment and revenue from patient consultations with different roles at the practice.

   In this study, we focus from an organisation's point of view and hence attempt to simplify the financial model in Figure 6.1 to Table 6.1. we assume every GP consultation is a 15-min slot and every nurse consultation is a 20-min slot. As nurse roles usually involve monitoring services (NZNO, 2016), we assume longer consultation time for the nurses. Knowing the workload would help the PHC centre to better manage their LTC patients, therefore, we assume that every consultation strictly follows these slot timings; which highlights that a GP can consult four patients and a nurse can consult three patients per hour. The PCA suggested to consider that every GP consultation will cost $40 for the patient and each nurse consultation costs $20, for a patient with these roles paid a salary of $100 (PayScale New Zealand, 2018) and $30 per hour (NZNO, 2016, p.9)

[2]This was a learning from the feedback through an informal meeting with the said role.

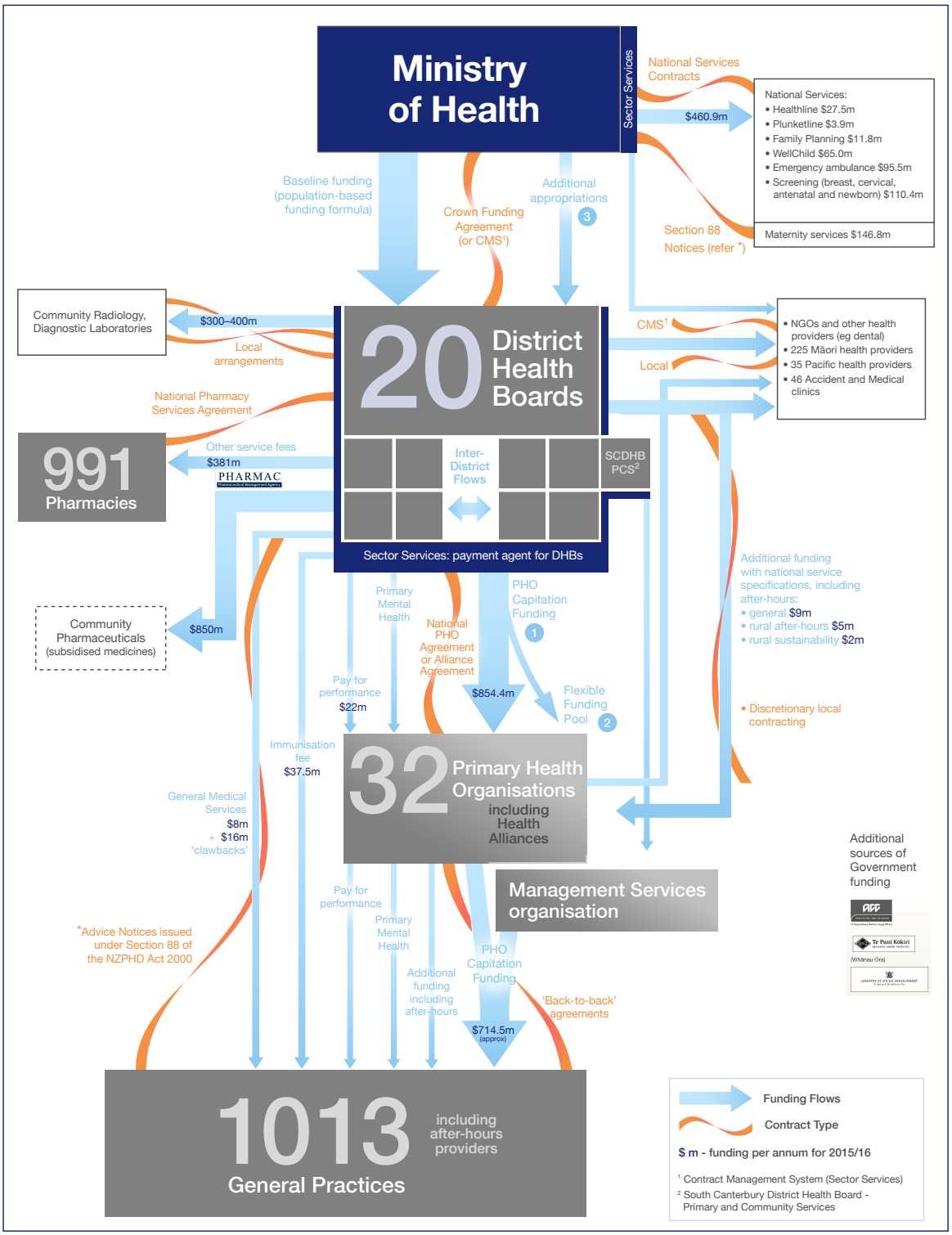


Figure 6.1: The New Zealand's Ministry of Health Funding model (Ministry of Health NZ, 2016e).

respectively. Table 6.1 summarises the net income from a GP and a nurse hour at a primary care centre. We also emphasise that these calculations are made simple for clarity, without taking into considerations the funding from the PHOs. Please note that we consider additional overhead only in case where we shift patients to nurses (we assume the nurses need to be provided with a consultation room to attend to these shifted patients, while the GPs continue to use the same resources).

Table 6.1: Income based on consultation.

Role	Revenue/hour	Salary/working hours	Total income per hour
GP	\$40 * 4 = \$160	\$100	\$60
Nurse	\$20 * 3 = \$60	\$30	\$30
Overhead			-\$25
Total	\$220	\$130	\$65

In the following section, we analyse a few what-if scenarios, with a comparison of our predicted workload (Predicted Workload in the graphs), how the LTC workload from the same patients was handled historically (the Historical Workload in the graphs) and the workload if the scenario is implemented. We consider their impacts including both the clinical and the financial risks involved for the year 2014. These analyses would help understand to what extent the organisations can reform to be proactive

- within their limited resources (discussed under Section 6.1), or
- with the changes in practice policies (discussed under Section 6.2), or
- when external factors impact the workload (discussed under Section 6.3), or
- when new models of care are considered for LTC patients (discussed under Section 6.4).

6.1 Resource-bound scenarios

The predicted workload impacts the resources at the practice. Usually, the number of resources available at a practice is limited. When the predicted workload is higher than the capacity of the practice, we say there is a *capacity mismatch*. The historical workload and predicted workload (number of LTC appointments per week in y-axis)

in Figure 6.2 clearly depicts this capacity mismatch problem. In our case, the capacity mismatch is a reflection of how the medical practice would need to address the (increased) workload if the practice chose to follow best practice guidelines for LTC patients. This also requires the medical practice to plan for either more resources or to shift the patients among the resources available, as discussed in the following sections. Please note here we are considering the capacity of *partner* GPs of the practice. Partner GPs of a practice are the GPs who have permanent contract with the practice and are unlikely to leave a practice. So, the practice registers LTC patients with them, although the patients can book appointment with any available GP. Other GP roles at the practice are considered as additional resources to meet the workload at the practice. The practice, whose data used in this study, has 5 GPs who have LTC patients registered with them.

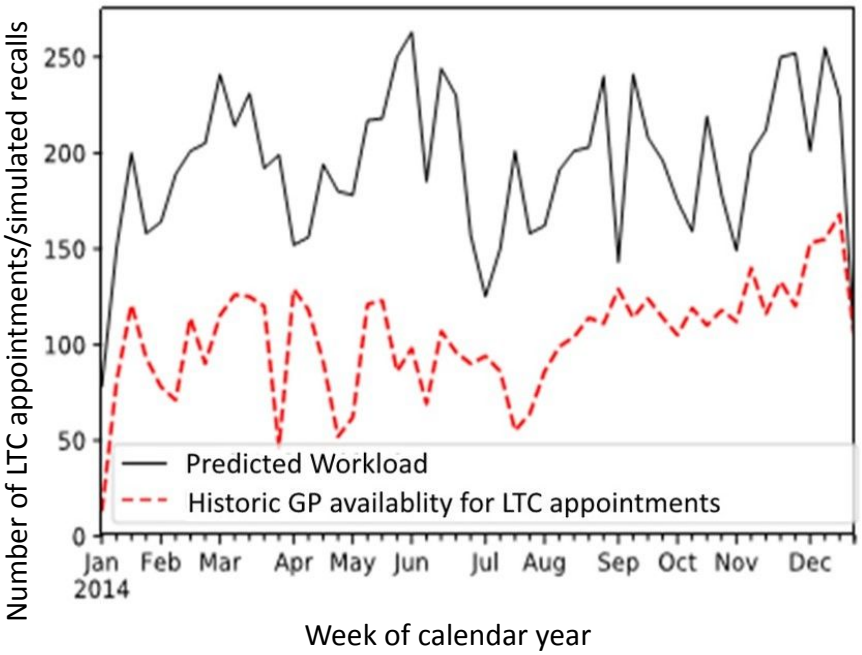


Figure 6.2: The capacity-mismatch problem.

A clinical risk analysis would say the practice should equip itself with a care provider capacity to address the excess 4 731 LTC appointments spread across the year 2014. Henceforth, in this section we discuss a few scenarios on the impact of proactive measures presented on the predicted workload.

6.1.1 Making GPs available for a specific number of LTC patients per working day

According to the dataset used in this study, the medical practice has five GPs who have LTC patients registered with them. These GPs are considered available for LTC patients; however historical availability of these GPs varies across weekdays. Heroman et al. (2012) argues that, under any circumstance, making the gap between the demand for care and the resource availability to zero would require over staffing and is not desirable. Therefore, here we enhance our rule-based model to have an additional capability to schedule recalls based on the total number of slots (i.e., the number of patient appointments a specific care provider can attend).

As highlighted (lines 8 and 9) in Listing 6.1, this rule accumulates the total number of slots available for the role "GP", and when there are more recalls scheduled, our rule randomly selects one recall (line 10 in the listing) and pushes them into the next working day (line 15). This continues until no more extra recalls are available (line 9 will then no longer hold true). In our case we set the `canAttend` field for all GPs as 10 (assuming the half of a weekday of available GPs is assigned to attend LTC patients), so if there were two GPs who attended LTC patients historically on a day in 2014, in our simulation this rule will assume the GPs can attend twenty LTC patients in total on that day. Figure 6.3 shows that imposing such a resource bound policy can help reduce the gap in capacity mismatch in attending LTC patients at the practice. This works because (as evident in the figure), historically `canAttend` per GP was less than 10. Assuming different values for such constraints can help the PHC to plan availability if more (or less) GPs will be available during a week.

Listing 6.1: The additional rule part needed to define the resource bound scenario.

```
1 rule "push recalls to next day if GPs non-available scene two"
2 salience 4700
3 when
4   $t:Today()
5   eval (workingday($t.today,Scenarios))
6   eval (Scenarios == 2)
7   $recall:Recall(recallDate.getTime() == $t.today)
8   $num_GPs_slots_available:  Number(intValue >= 0) from accumulate ($role:
     CareProviderRole(roleName =="GP",dateOfAppointment ==
     convertToDate($t.today),$total_slots:canAttend), sum($total_slots))
9   $recalls:  List(size>=0, size > $num_GPs_slots_available.intValue()) from
```

```
    collect (Recall(recallDate.getTime() == $t.today))
  $randomRecall:Recall() from $recalls
then
  Recall $newrecall = new Recall();
  $newrecall.patientId = $randomRecall.patientId;
  $newrecall.recallInFreq = $randomRecall.recallInFreq;
  $newrecall.recallDate = java.sql.Date.valueOf($randomRecall.
   recallDate.toLocalDate().plusDays(1));
  $newrecall.recallComment = "pushed recall due to GP
   unavailability scene two";
  $newrecall.consulted = "GP";
  insert($newrecall);
  retract($randomRecall);
end
```

For now, let us consider that these appointments are attended by GPs only. The comparison of net income from the historic GP workload, if the predicted workload is attended by GPs and if GPs are made available for 10 LTC patients per working day is given in Table 6.2. Also, adding those constraints means that these GPs attend those scheduled patients only. As depicted in the table, ideally if the practice attends all the predicted LTC patient appointments (although it is not desirable (Heroman et al., 2012)), the net income would bring an additional $70,965. If the practice makes the available GPs to attend 10 LTC patients per working day, would bring an additional income of $40,140 i.e., an additional income equivalent to (approximately) 50% of the current net income.

Table 6.2: Comparison of income in three different cases of workload management

Scenario	**Number of appointments**	**Net Income**
Historical number of appointments	5 454	$81,810
Predicted number of appointments	10 185	$152,775
Making the available GPs to attend 10 LTC patients per working day	8 130	$121,950

As depicted in Figure 6.3, the comparison between predicted workload and this what-if scenario also shows that there will still be unmet demand for care. We remind the reader that these what-if scenarios highlight the impact of different scenarios on the

predicted workload. Therefore, the organisational manager (with a basic knowledge about the rules in this tool) can use this tool to analyse the availability of the number of GPs as well as the number of LTC patients needing to be assigned to those available GPs. We acknowledge that having more LTC patient slots makes the GPs less available for the non-LTC cases. We argue that having an understanding of how long (number of time slots) the partner GPs are unavailable to attend non-LTC patients, the practice will be able to equip with other non-partner GPs or nurse roles to attend the non-LTC cases accordingly.

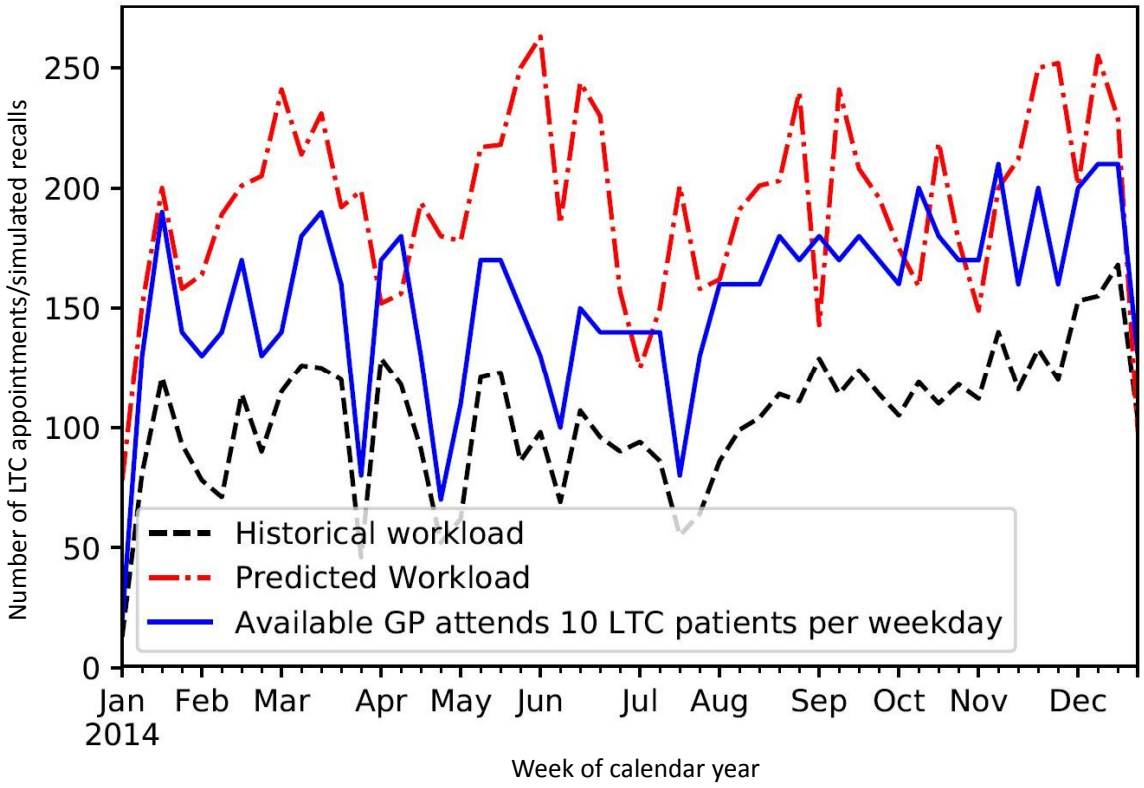


Figure 6.3: Making GPs available for 10 LTC patients per working day.

6.1.2 Make nurse appointments available

Nurse appointments are not new in primary care, for instance, there are nurse clinics for severe CKD patients (Potts et al., 2011). However, usually, clinical conditions and prevalence of the condition among other factors determine the funding model for managing that condition, which in turn impacts implementation of nurse appointments. Moreover, a PHC practice receives funding based on the enrolled patients and the GP

services they use (Ministry of Health NZ, 2014). Although there are initiatives such as the Practice Nurse Subsidiary by the Ministry of Health NZ (2017a), most general practices provide GP services in preference to nurse services because of the funding model. This makes the clinical managers sceptical about what would happen to the income if their patients visited nurses instead of GPs. In this section, we argue that nurse clinics will not only reduce clinical risks but also will help to bring more revenue to the clinic.

Clinically, less severe patients require timely health assessments to manage and prevent LTC associated complexities, and motivation to self-manage their conditions. Having longer appointments with nurses would help to build a good relation between the nurses and their patients, which in turn, motivates patients to better self-manage their health (Busetto et al., 2017). In order to have more generalisable rules, patient classifications are updated with an attribute to denote the *severity* of a condition in them. In this way, new patients can be stratified at the time of diagnosis and added with a care plan that depicts the severity of the condition in them.

Let us consider a scenario where the less severe[3] patients are attended by the nurses. We remind the reader that we assume that the various roles are competent to attend these shifted patients.

Our rules can, then along with other conditions and one more condition (given in line 8 in Listing 6.2) select these less severe patients and assign them to be consulted by a nurse (line 14 in the listing).

Listing 6.2: The additional rule part added to shift less severe patients to a nurse.

```
1 rule "Push low severe patient recalls to Nurse"
2 salience 4700
3 when
4 $t:Today()
5 eval (workingday($t.today,Scenarios))
6 eval (Scenarios == 3)
7 $recalltoday:Recall(recallDate.getTime() >= $t.today,consulted
    != "Nurse")
8 $cp:  CarePlan(patientId == $recalltoday.patientId,severity == "low")
9 then
```

[3]Severity of LTCs in a patient is determined by the care plans applicable to them. For this what-if scenario, we consider the severity of diabetes, depicted in the diabetic risk score of a patient, and severity of CKD depicted in the CKD stage for a patients.

```
10 Recall $newrecall = new Recall();
11 $newrecall.patientId = $recalltoday.patientId;
12 $newrecall.recallInFreq =$recalltoday.recallInFreq;
13 $newrecall.recallDate = $recalltoday.recallDate;
14 $newrecall.consulted = "Nurse";
15 $newrecall.recallComment = "This is a Nurse consultation for
     low severe patients"
16 insert($newrecall);
17 retract($recalltoday);
18 end
```

Figure 6.4 shows the predicted workload and the workload for GPs and nurses if less severe patients of our cohort are shifted to nurses. In our case, we assume that more nurse slots will be allocated to attend the patients shifted to them. This also means the GP slots which were scheduled for the less severe patients are now freed (the GPs line depicts this reduced workload) and those slots are available for other patients.

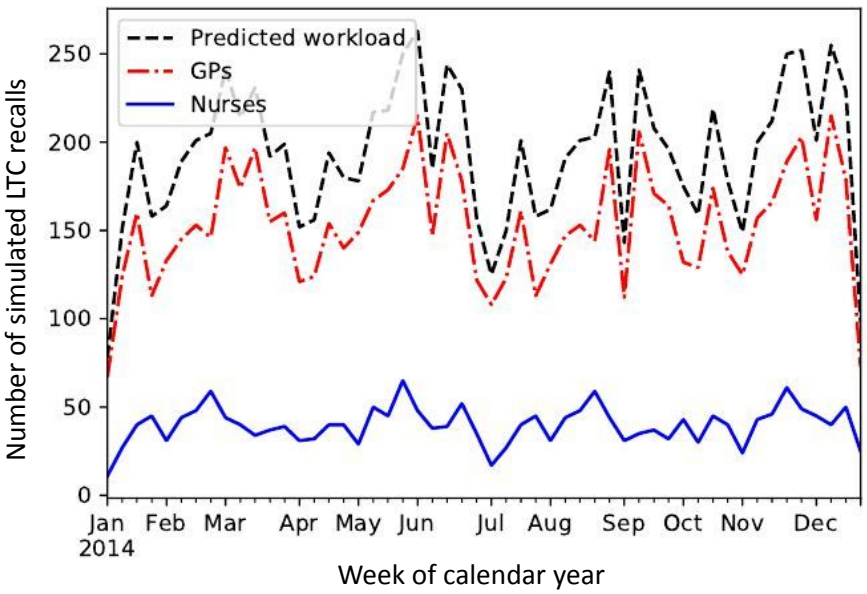


Figure 6.4: Impact of shifting less severe patients to nurses on the predicted workload.

Financially, shifting patients to nurses would bring more revenue (i) as more GP slots are available for other patients, and (ii) nurse consultations, too, bring revenue.

Based on the financial model at the beginning of this chapter and Figure 6.4, if, out of the predicted workload (10 185 appointments in 2014), the less severe patients are attended by nurses (out of 10 185, 2 098 appointments can be shifted to nurses). This

will free 2 098 GP slots for more patients as well as providing more income through nurse consultations. Adding the GP consultations and nurse consultations with the GP slots freed to be available for other patients would give $173,820. However, assuming an overhead of $25 per hour for nurse consultations, will decrease the income by $17,500. Considering the income based on Table 6.1, subtracting that overhead from the income would give a total of $156,320. Compared to the current practice, this would bring an additional income of $74,510. This is summarised in Table 6.3.

Table 6.3: Financial risk analysis for including nurse consultations for low-severity LTC patients.

Role	**Number of patients**	**Number of role hours needed**	**Total income**
GP	8 087	2 022	$121,320
Nurse	2 098	700	$21,000
GP fill slots	2 098	525	$31,500
Practice overhead	2 098	700	-$17,500
Scenario net income			$156,320
Current practice income			$81,810

In a similar way, the model can be used to analyse shifting patients from GPs to other care provider roles and can also be scaled to choose the cohort of patients to be shifted. By scaling, we mean to increase the number of patients to shift to other care provider roles. For example, instead of less severe patients, considering one visit of every LTC patient to be attended by nurses, in our case, this would account for 4 190 patient appointments shifted to nurses. The following section describes the impacts of this what-if scenario.

6.1.3 Every LTC patient with multiple visits with the practice visits a nurse once a year

One another proactive way to make GPs available for other patients is to shift one visit among the multiple visits of a patient to a nurse. This will help to maintain continuity of care as well as the nurses becoming familiar with their patients' needs (McPhail, 2016). Financially, this shift will have a similar impact as discussed in the previous

section.

The best practice followed for this work emphasises to recall an LTC patient at least every six months, which means there are at least two visits per LTC patient in a year. If one of these LTC related visits are shifted to nurses, then in our case 4 190 visits will be shifted to nurses. This will account for freeing 4 190 GP slots and making them available for more patients. This will produce a total income of $159,805 as given in Table 6.4, which means an additional income of $77,995. In Listing 6.3, the rule considers that all the scheduling is initially made and one visit among the many visits scheduled for a patient is assigned to a nurse. Hence, we add a constraint that this rule is evaluated at the end of the simulation (line 5) and that it should be evaluated last among the other rules activated (line 2). We also ensure that this rule should not cause the rule engine to re-evaluate the rule set because of changes in the consequence of this rule (line 3). We also check that the same patient is not assigned to a nurse twice (line 8 in the listing), however, the practice can change these rules to fit their requirements. The model can be used to analyse other similar scenarios as well.

Listing 6.3: The additional rule part added to shift one visit among multiple visits of a patient to a nurse.

```
1 rule "Shift one recall from multiple recalls of a patient to a
    Nurse"
2 Saliance -1
3 no loop true
4 when
5 $t:Today(today == SimulationEndDate)
6 eval(Scenarios == 4)
7 $recall: Recall(recallDate.getTime() <= SimulationEndDate)
8 not (exists Recall(patientId==$recall.patientId,consulted == "Nurse"))
9 $multipleRecalls:  List(size >1) from collect
    (Recall(patientId==$recall.patientId,recallDate.getTime()<=SimulationEndDate)
    )
10 $randomRecall:  Recall() from $multipleRecalls
11 then
12 Recall $newrecall = new Recall();
13 $newrecall.patientId = $randomRecall.patientId;
14 $newrecall.recallInFreq = $randomRecall.recallInFreq;
15 $newrecall.recallComment = "This is a random recall among many
    recalls to a Nurse"
```

```
16 $newrecall.recallDate = $randomRecall.recallDate;
17 $newrecall.consulted = "Nurse";
18 insert($newrecall);
19 retract($randomRecall);
20 end
```

Table 6.4: Financial risk analysis if shifting one visit among many visits of a patient to a nurse.

Role	**Number of patients**	**Number of role hours needed**	**Total income**
GP	5 995	1 499	$89,940
Nurse	4 190	1 397	$41,910
GP-fill appointments	4 190	1 048	$62,880
Overhead	4 190	1 397	-$34,925
Scenario net income			$159,805
Current practice income			$81,810

6.2 Change in practice policy

In this section we discuss what the predicted workload would be if there are changes in availability of GPs at the practice. One such scenario is when the practice decides to increase its GP and nurse availability for their LTC patients, for example, a Saturday practice for LTC patients. We take this opportunity to highlight that this scenario is to analyse the impact of increasing GP and nurse slots during a week. We also highlight that we are considering Saturday as a (normal) working day and do not consider it as overtime or a weekend pay for the GPs and nurses available on Saturdays.

6.2.1 Saturday practice for LTC patients

Let us assume that the practice decides to increase the availability of LTC appointment slots, with a GP (i.e., 20 more GP slots) and 3 nurses (i.e., 40 more nurse slots) on a Saturday for their LTC patients. The flexibility in our rules allows us to avoid including Saturdays as part of weekends, and to have the capability to add the additional number of slots for a specific day. Listing 6.5 highlights this capability. Please note here we

have used `totalPatientsAttendedActualData` (line 6) so that this decision on the number of patients to be assigned to other roles can also come from the actual data. When there are more patients to be assigned to care provider roles than the total number of slots available (line 7), a recall (line 8) is pushed to the next day (line 13). This continues until no more recalls need to be pushed to another day (i.e., the line 7 no longer hold true). Our rule base also considers holiday list based on scenarios (value 7 on line 4 here means: Saturday also should be considered as a working day). The solid line in Figure 6.5 below shows that the workload is managed better with an extra working day (Saturday in this case) for the LTC patients.

Listing 6.4: The additional rule part added to add additional slots for its LTC patients for one more day at the practice.

```
1  when
2  $t:Today()
3  eval (workingday($t.today,Scenarios))
4  eval (Scenarios == 7)
5  $recall:Recall(recallDate.getTime() == $t.today)
6  $no_GPs_slots_available:  Number(intValue >= 0) from accumulate ($role:
       CareProviderRole(dateOfAppointment ==
       convertToDate($t.today),$total_slots:totalPatientsAttendedActualData),
       sum($total_slots))
7  $recalls:  List(size>=0, size > $no_GPs_slots_available.intValue()) from
       collect (Recall(recallDate.getTime() == $t.today))
8  $randomRecall:Recall() from $recalls
9  then
10 Recall $newrecall = new Recall();
11 $newrecall.patientId = $randomRecall.patientId;
12 $newrecall.recallInFreq = $randomRecall.recallInFreq;
13 $newrecall.recallDate = java.sql.Date.valueOf($randomRecall.
       recallDate.toLocalDate().plusDays(1));
14 $newrecall.recallComment = "pushed recall due to GP
       unavailability scene seven";
15 $newrecall.consulted = "pushed from sat_practice";
16 insert($newrecall);
17 retract($randomRecall);
18 end
```

Given a GP can attend 20 patients per Saturday, except three public holidays that fell on a Saturday in 2014, there are 49 working Saturdays at the practice. This accounts for a total of 980 GP appointments in the year 2014. Similarly, there are 1 960 nurse appointments. Deducting the overhead of running this extra working day at the practice, the net income will be $17,940. Table 6.5 summarises these financial aspects.

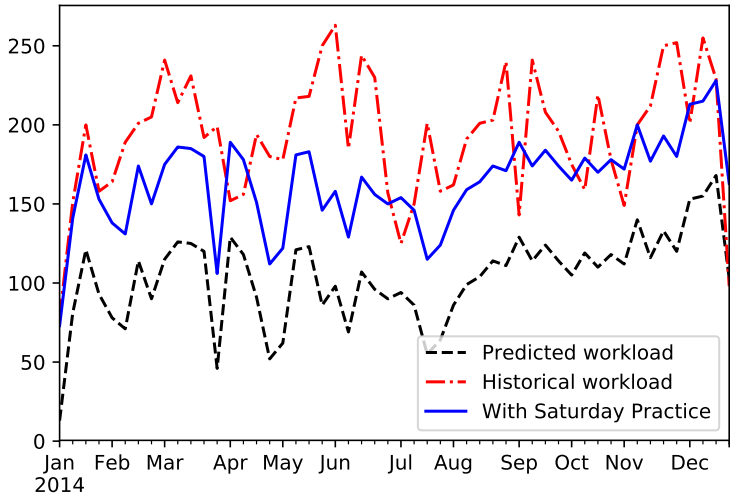


Figure 6.5: Impact of having Saturday practice.

Table 6.5: Financial risk analysis if practice dedicate a GP and 3 nurses for its LTC patients for one more day each week at the practice

Role	**Income per hour**	**No: of patients**	**Total income**
GP	$60	980	$14,700
Nurse	$30	1960	$19,620
Overhead	-$25	1960	-$16,350
Net income			$17,970

6.2.2 All patients should have a screening for CVD once a year.

Knowing how many patients are affected by a new health scheme can help the practice decide the right role to attend this additional workload. Let us assume that the Ministry

of Health NZ, has announced a new health strategy to have CVD[4] screening for all LTC patients at the practice, which means (in our case) there is an extra workload from 4 190 patients in the coming year. Please remember that this is an additional workload to the predicted workload. With a 15-min slot for GPs, this additional workload will require 1 048 GP hours. However, the main competencies of nurses include monitoring and patient education (NZNO, 2016). Hence, assuming this workload is handled by nurses, it would require 1 397 nurse slots which will bring $41,910. With a total overhead $34,925, the net income would be $6,985. Having nurses attending this additional workload, the organisation can manage the predicted workload as well the additional workload due to the (new) health scheme.

6.3 Impact of external factors on the workload

At times, the workload at the practice is affected by factors external to the organisation. For example, during the winter time there is a high probability of an increase in workload due to an influenza outbreak, although the timing cannot be forecast. Even though the practice could equip itself with more GPs and nurses, the workload due to LTC management will have to be managed separately.

Consider the following scenario: a flu outbreak happens in the first week of May and the workload from LTC patients for the next two weeks has to be managed. Listing 6.5 shows one possible way to tackle this workload. All the three-monthly recalled patients (line 8) are prescribed their medications over the counter (or a Nurse see line 17) and scheduled to have a recall in another three months (line 16 in the listing) and thus avoid an appointment with a GP during the flu period. And, the patients who are on a six-monthly review (line 29) will have their follow-up booked with a nurse (see line 37) in another three months (line 36). According to our PCA, this measure will address clinical risks for LTC patients, and the six-monthly recalled patients are not left on their own for a long period (another six months if they are not able to make an appointment during the influenza period). Please note that the rules (lines 9 and 30) select a random recall among the three-monthly and six-monthly recalled patients. However, we can provide a constraint to consider a subset of n patients (by modifying List size equals some value n as List (size = n) instead of List (size > 0) where it selects *all* patients).

[4]Cardio-Vascular Disease

Listing 6.5: The additional rule part added to handle external factors (influenza outbreak).

```
1  rule "manage scenario five 3 monthly recalls"
2  salience 7000
3  when
4  $t:Today(today > FluStartDate && today <FluEndDate)
5  eval (workingday($t.today,Scenarios))
6  eval (Scenarios == 5)
7  Recall(recallDate.getTime() == $t.today)
8  $3monthlyRecalls:  List(size >0) from collect (Recall(recallDate.getTime()
       == $t.today,recallInFreq == 3 ) )
9  $randomRecall:Recall() from $3monthlyRecalls
10 then
11 System.out.println("**********************in rule : Date = " +
       convertToDate($t.today))
12 Recall $newrecall = new Recall();
13 $newrecall.patientId = $randomRecall.patientId;
14 $newrecall.recallInFreq = $randomRecall.recallInFreq;
15 $newrecall.recallComment = "Recall skipping a regular 3 monthly
       "
16 $newrecall.recallDate =
       java.sql.Date.valueOf($randomRecall.recallDate.toLocalDate().plusMonths(3));
17 $newrecall.consulted = "Nurse";
18 insert($newrecall);
19 retract($randomRecall);
20 end
21
22 rule "manage scenario five 6 monthly recalls"
23 salience 7000
24 when
25 $t:Today(today > FluStartDate && today <FluEndDate)
26 eval (workingday($t.today,Scenarios))
27 eval (Scenarios == 5)
28 Recall(recallDate.getTime() == $t.today)
29 $6monthlyRecalls:  List(size >0) from collect (Recall(recallDate.getTime()
       == $t.today,recallInFreq == 6 ))
30 $randomRecall:Recall() from $6monthlyRecalls
```

```
then
Recall $newrecall = new Recall();
$newrecall.patientId = $randomRecall.patientId;
$newrecall.recallInFreq = $randomRecall.recallInFreq;
$newrecall.recallComment = "Recall making a 3 monthly of a
    regular 6 monthly"
$newrecall.recallDate =
    java.sql.Date.valueOf($randomRecall.recallDate.toLocalDate().plusMonths(3));
$newrecall.consulted = "Nurse";
insert($newrecall);
retract($randomRecall);
end
```

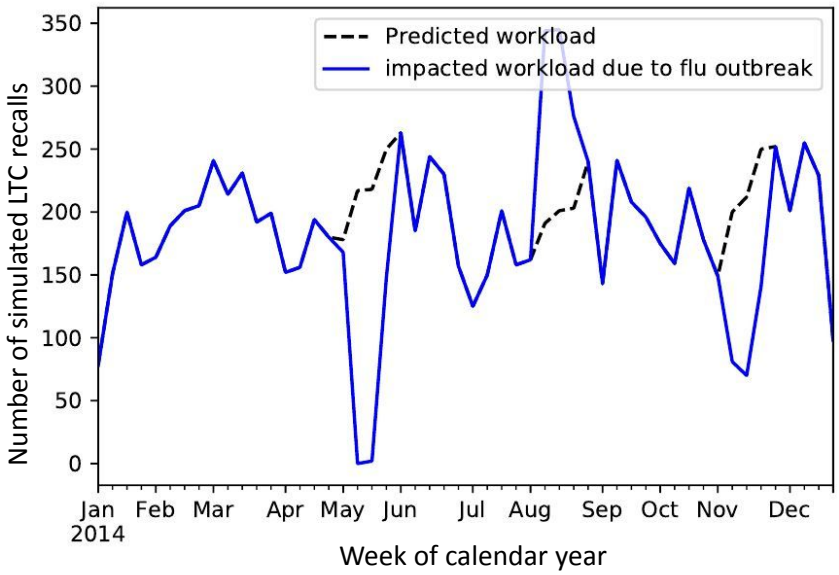


Figure 6.6: Requirement to have fewer LTC appointments during influenza outbreak will have adverse impact on predicted workload.

Our model will predict this proactive action to accommodate the incoming acute cases with a drop in the number of LTC appointments in the first week of May and have an increased workload in the last week of July and the first two weeks of August (refer to Figure 6.6). This will also have an impact on the workload in late-October early-November weeks where only the regular three monthly patients are recalled. Being informed of this impact, the organisation can plan to make more GPs or nurses available during that period. On the other hand, rather than pushing the six-monthly reviews to

three months, the organisation can also evenly distribute the workload in the following weeks of influenza, if the organisation can make GPs and nurses available accordingly.

6.4 Implementing new models of care

One of the major problems with the models of care discussed in Chapter 2 is that they generally lack the ability to adapt to the changing care needs of their LTC population (Bodenheimer et al., 2002a). In this section, we show our model is also able to incorporate new models of care in the form of new care plans for a population. For instance, Ministry of Health NZ (2017b) defines Care Plus as an initiative to support people with high health needs, including those related to LTCs. It is stated that eventually the High Use Health Card (HUHC) patients will be assessed to move to Care Plus. Moreover, the overall funding for Care Plus patients depends on number of Care Plus patients and the services used by these patients.

Listing 6.6 considers a scenario where the organisation plans to recall HUHC patients (we refer HUHC patients as `carePlus` patients, hence using carePlus == true in line 6) every three months (line 11). This recall for three months is applied to patients who are on recall frequencies greater than three months (line 7). Please recall from Chapter 4 that individual recall periods for each LTC in a patient are recorded as the `IndividualRecallDetail` facts for the patient, which are accumulated to decide the recall date a patient (see Section 5.2.2). Therefore, in order to add this new model of care, we add an `IndividualRecallDetail` fact (line 9) which recalls a patient in three months (line 11) on a care plan coded "careplus" (line 12). This will obviously increase the workload as shown in Figure 6.7.

Listing 6.6: The additional rule part added to adapt new models of care.

```
1 rule "Insert CarePlus care plan"
2 salience 7000
3 when
4 $t:Today(today == SimulationStartDate)
5 eval(Scenarios == 6)
6 $p:Patient(carePlus == true)
7 not (exists IndividualRecallDetail(patientId == $p.patientId,
     recallInFreq <= 3))
8 then
9 IndividualRecallDetail $newindrecall = new
```

```
    IndividualRecallDetail();
$newindrecall.patientId = $p.patientId;
$newindrecall.recallInFreq = 3;
$newindrecall.carePlanCode = "careplus"
insert($newindrecall);
end
```

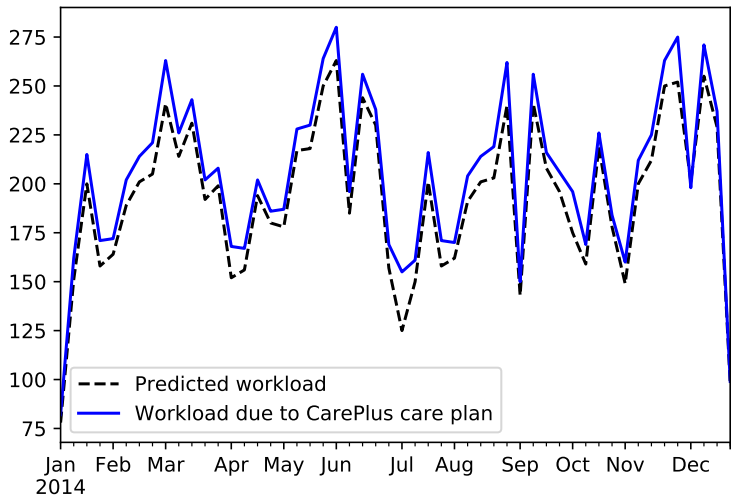


Figure 6.7: Increase in workload due to a new model of care.

The Ministry of Health Care Plus scheme has much more impacts than the financial analysis presented in this section. Please bear in mind that the scenario shows the capability of our model to add new plans-of-care for a patient, and the model can predict and analyse the workload if that plan-of-care is present for a patient. On the financial side, there will be an additional 744 appointments in 2014. Given that the funding is for GP appointments, assuming these appointments are consulted by GPs this will account for additional income of $11,160.

Table 6.6 summarises the financial analysis of the various scenarios discussed in this chapter (except the impact of influenza). These scenarios are for the year 2014. The GP funding models are complex and hence the calculations are kept simple (for clarity) to consider the income from patient consultations and costs of employing the GPs and nurses. The net income shown in the table considers the overhead cost as well. The figures used are approximate values that are closely related to the values from various sources such as PayScale New Zealand (2018) and New Zealand Nurses Organisation (NZNO, 2016). As presented above, with these scenarios being implemented as rules, the practice has the flexibility to add more rules to analyse different scenarios applicable

to their context and practice policies.

Table 6.6: Summary of financial analysis of the various scenarios discussed in this chapter.

Scenario	**Total number of appointments**	**Net Income**	**Additional income**
Historical number of LTC appointments	5 454 (GP)	$81,810	$0
Predicted number of LTC appointments	10 185 (GP)	$152,775	$70,965
Available GPs attend 10 patients per working day	8 130 (GP)	$121,950	$40,140
Less severe patients shifted to nurses	8 087 (GP) + 2 098 (nurse) + 2 098 (GP-fill slots)	$156,320	$74,510
One visit of every LTC patient shifted to nurse	4 190 (nurse) + 5 995 (GP)	$159,805	$77,995
Increase GP and nurse availability in a week	980 (GP) + 1 960 (nurse)	$99,780	$99,780
CVD screening for all patients	4 190 (nurse)	$88,765	$88,765
Implementing new model of care	744 (GP)	$92,970	$92,970

In summary, unavailability of GPs and growing demand for LTC care emphasise the need for a tool to help PHC practices manage their LTC workload better. We developed a workload prediction model that, given the LTC patient information, predicts the upcoming workload as a volume of number of appointments from its LTC patients. Knowing that any new policy or procedure can have clinical and financial impacts, we analysed a few what-if scenarios in these aspects. We showed that our model can be used to analyse different scenarios, including implementation of new models of care. Having shown the strength of our model to analyse the alternatives at a PHC centre, we claim that our model can help practices to shift from a patient-initiated reactive system to a proactive systematized LTC care system.

As part of DSRM methodology, we presented these scenarios to the members of

Otago and Southern PHO including CEO, CIO and general manager of WellSouth Primary Health Network, a representative of nursing organisation, and the Medical Director Strategy, Primary and Community of Southern DHB. They found the work to be useful and informative. A major feedback was that the practice managers are aware that there are "swings" in the availability of GPs during a year but "we are not aware of how much these swings impacts and the variations are on the workload of the organisation as a whole. That is very interesting...". Another feedback was along the lines of having such a tool will improve the ability of a practice to have an informed decision about its "workforce development" (this justifies our RQ2 and RQ3 in Section 1.2). They also acknowledge that stratifying patients based on their LTC needs is essential, and that they currently are working on a home care model for patients with very high need for LTC care.

Chapter 7

Using Bayesian inference in the Workload Prediction Model

A care pathway (CP) suggests the time line for various care actions to be followed to manage a specific LTC in a patient (NICE, 1999). This includes patient recalls[1]. However, in a patient database, although it captures the care actions and their timestamps, it is not explicit which care pathway was applied or the reason for these care actions. Therefore, (revisiting Chapter 4) in order to identify LTC visits we consider the prescriptions of active LTC medications of our patients. We assume that medications are prescribed following a 90-day repeat cycle for LTC medications as is required in New Zealand, and patients are implicitly reminded to visit or at least contact the practice when they need to get their prescription reissued.

In other words, our simulation model assumes that patient visits will be driven by best practice rules for recalling patients, but the dates of those visits will align with the cycle of 90-day prescription periods. In our predicted workload, in rigidly following the best practice guidelines, we assumed that the patients visited on their 90-day medication run-out day. However, we observed periodic peaks (specifically in the visit pattern of the three-monthly recalled patients) that are not observed in the historic pattern of visits for these patients (see Figure 7.1). We suspect that it could be due to using the most recent visit in year 2013 (the line labelled as visits in 2013 that predict the recalls) to simulate subsequent visits for these patients. We, also, have observed that in historic data (as highlighted in Section 5.3) some patients have more frequent visits than our rules would predict. This could be because, for patients

[1]Please remember that, here, recalls are the follow-up appointments required to manage a specific LTC in a patient.

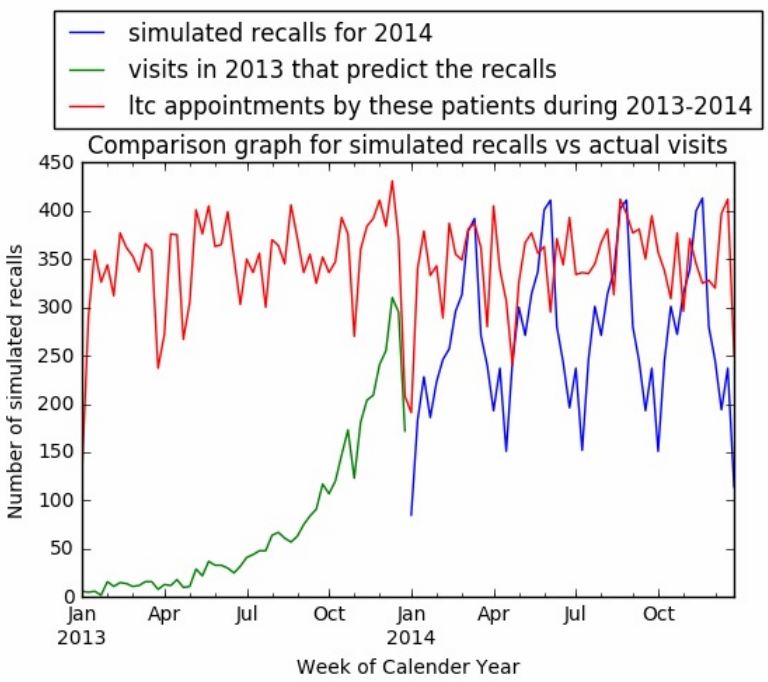


Figure 7.1: Initial results of simulation for 3-monthly recalled patients.

with recall periods greater than 90 days, some reissuing of prescriptions would happen over the phone between their scheduled visits, or a few visits may not be related to medication running out, e.g., those due to acute exacerbation of their LTCs.

In Chapter 4, we also showed that for the case of patients with CKD only, when the most recent visit of a patient in the year prior to our simulation period is used to initiate our simulation, this resulted in an accumulation of simulated visits towards the end of the year (see Section 4.6.2).

These issues highlight two shortcomings of our current approach:

- The most recent visit prior to the simulation may not be representative of their periodic visit behaviour aligned with medication prescription cycles. As presented in Section 4.6.2, we used a random visit among multiple visits in the previous year to break the periodicity from choosing the most recent visit. However, the randomly chosen visit also may not be a periodic visit due to medication running out.

- Using a single most recent historic visit date means we have no information about the variance of a patient's visit dates relative to when their medication is due to run out. Instead we previously assumed that patients will visit on the date their medications will run out (adjusted to avoid weekends and public holidays). This

causes unrealistic repeating pattens in the predicted workload.

To address these problem, in this chapter we propose the use of Bayesian inference on a sequence of historic visit dates to learn a probability distribution for each patient's recurring 90-day cycle of repeat prescriptions relative to a given historic reference date. This has two benefits:

1. We get a better estimate of when each patient's medication will run out, based on more data. We also learn a parameter representing their likelihood of deviating from this date.

2. Rather than assuming in our simulation that patients will visit in some multiple of 90 days since their last visit (with the multiplicative factor depending on their recall period), during the simulation we can sample from a learned probability distribution representing each patient's likelihood of visiting on a certain date relative to the date the medication is due to run out. This gives a more natural randomness to the visit dates.

To use Bayesian inference, we define a parameterised statistical model, and learn the conditional probability of these model variables given the observed data (Salvatier et al., 2016). The belief we have about the probability of a model parameter is defined as a *prior* distribution of that parameter. Thus, Bayesian inference uses the information provided by observed data, formally the *likelihood*, to update a *prior* state of beliefs to become a *posterior* state of beliefs about a (set of) parameter(s). In other words, the probability distribution that is learned from the parameters' *prior* distribution and the *likelihood* is the *posterior distribution*.

As discussed by Wiecki (2015), given a model that predicts the likelihood P($\mathbf{x}$ | θ) of observed data $\mathbf{x}$, (where θ is a vector of model parameters), Bayes theorem (see the equation below) can be used to find a probability distribution over the model parameters, P(θ).

$$P(\theta \mid \mathbf{x}) = \frac{P(\mathbf{x} \mid \theta).P(\theta)}{P(\mathbf{x})}$$

where $P(\mathbf{x}) = \int_\theta P(\mathbf{x} \mid \theta)\ d\theta$ is an integral over all possible values of parameters (θ).

When the prior distribution and likelihood are "conjugate"(belong to the same family of probability distribution), the integral in the denominator can be determined analytically. However, in general, this is not possible, i.e., the integral in the denominator of Bayes Theorem cannot be computed in a closed (formulaic) way. Hence,

the Markov Chain Monte-Carlo (MCMC) process is commonly applied for inferring posterior distributions in Bayesian inference.

MCMC generates a large set of samples from the posterior probability distribution defined in a model. The random samples are generated by a special sequential process where each random sample is used as a stepping stone to generate the next random sample producing a "chain". A special property of the chain is that, while each new sample depends on the one before it, new samples do not depend on any samples before the previous one (this is the "Markov" property). The sampling techniques (e.g., Gibbs sampling) used in MCMC aim to ensure that these chains contain sample values in proportion to the probability distribution being learned (the posterior distribution in our case). Thus, eventually, the samples MCMC generates would appear as if they are coming from our posterior distribution. We can then use that sample of posterior probability to obtain an estimate of each patient's parameters of interest.

We sought statistical advice from the Department of Statistics and Mathematics of our university. In response to an initial model of ours, they suggested to use two parameters: one to represent the prior belief about the dates of medication running out for each patient (denoted as *medsRunOut*), and another to represent the likelihood of variation of visit from this medication run out day (denoted as *alpha*). In the following section, we describe our Bayesian inference model (developed based on this advice) in a mathematical notation (please refer to Appendix A.12 for its implementation using the PyMC3[2] library for the Python programming).

7.1 Model Parameters

We model the patient's cycles of medication running out as a series of 90-day intervals from a selected historical reference date (referred to as the Bayesian start date). Our model considers that a patient's medication will first run out on some day between 0 and 89 days (inclusive) starting from the initial reference date. Initially, we have no information about which day this is for each patient, so our prior distribution for this variable, medsRunOut, is a Discrete Uniform distribution over the interval [0,89].
$medsRunOut \sim DiscreteUniform(0, 89)$

As described before, medsRunOut denotes a day in the 90-day interval when medications would run out for a patient. For a given medsRunOut day, the patient visit day can vary, i.e., the patient may visit early, or later than the medsRunOut day for that

[2]More on PyMC3 at https://docs.pymc.io/

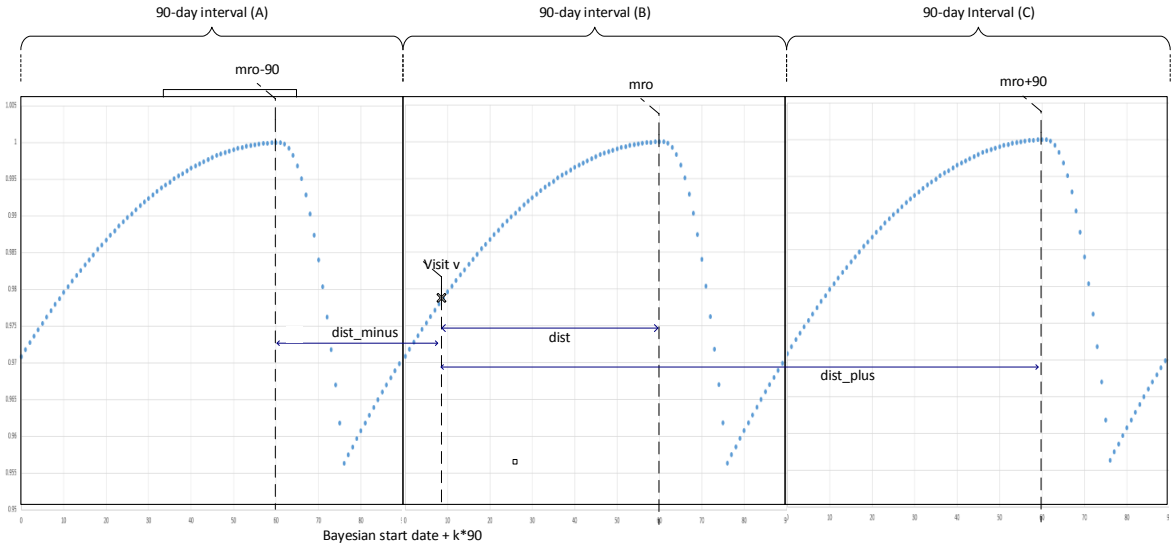


Figure 7.2: Likelihood of a visit

patient. We now consider the likelihood of a patient visit on a given day if medsRunOut is known. We represent an observed visit day as $(visit\ day - reference\ date)\ mod\ 90$. For instance (refer to Figure 7.2), consider a patient visit (Visit v) in interval B, whose medsRunOut value is 60 i.e., the patient's medication would run out in the 60th day in that 90-day interval. The figure shows three 90-day intervals namely intervals A, B and C. Having observed a patient visit (Visit v) in the Interval B, the visit can be

1. late with respect to the interval A;
2. prompt within the interval B;
3. early with respect to interval C (if Visit v was on the right side of mro in Interval B);

The likelihood of a visit on a given day is then defined using a function that scales the likelihood of a visit as an exponential function of the distance of a date from the medsRunOut date:

$$f(day) = e^{-alpha\ *\ distance\_between(day,\ medsRunOut)}$$

where the distance_between function[3] is defined in Algorithm 2. The function in Algorithm 2 implements the likelihood defined by considering the minimum distance between a visit date v and the current, previous and next medsRunOut (mro) dates

[3]Note that this function is not normalised so that its values sum to 1 over the integers [0,89], but this not required by the PyMC3 Python library used to implement Bayesian inference using MCMC.

Result: The distance v from mro in terms of 90-day interval

$dist\_minus = v - (mro - 90)$;

$dist = v - mro$;

$dist\_plus = v - (mro + 90)$;

$skew\_factor = 5$;

$value\_near\_mro\_minus = (skew\_factor^{sgn(dist\_minus)} * dist\_minus^2)$;

$value\_near\_mro = (skew\_factor^{sgn(dist)}) * dist^2$;

$value\_near\_mro\_plus = (skew\_factor^{sgn(dist\_plus)}) * dist\_plus^2$;

$min\_val = \min(value\_near\_mro\_minus, value\_near\_mro, value\_near\_mro\_plus)$;

return min_val;

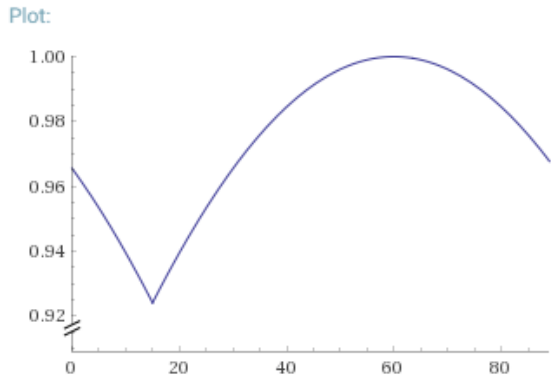
Algorithm 2: The distance_between function.

(i.e., dist, dist_minus, and dist_plus distances, respectively, in Figure 7.2). The function squares the distances to increase the likelihood of visits close to the medsRunOut date. The *sgn (i)* denotes a function that returns 1 if i is positive, 0 if i is zero, and -1 if i is negative. The skew factor is discussed below.

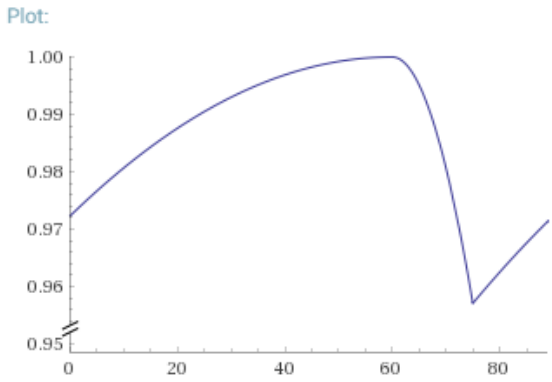
The patient visits may vary from the exact day of medications running out. Although we do not know how much this variation is, we assume that it is small (we assumed a three-week window in our LTC definition in Section 4.4.1). The scaling parameter (*alpha*) in the likelihood function f controls the amount that patient visits can deviate from the medsRunOut date. As a parameter of our model, *alpha* must have a prior distribution. Based on statistical advice, we use a weakly informative prior that is a t-distribution with a small-moderate variation (λ =0.04) with heavy tails (a degree of freedom 3) (Gelman, 2006; Gelman et al., 2008). In order to consider only positive values, the prior for *alpha* uses a HalfStudentT distribution.

$$alpha \sim HalfStudentT(\lambda = 0.04, \nu = 3)$$

We also assume that the likelihood for a patient to visit n number of days before medsRunOut is higher than n number of days after medsRunOut. Hence, in the distance function defined in Algorithm 2 we use a skew factor which is used to manipulate this increased likelihood of a patient visit before medsRunOut. Using a skew factor of 5, the likelihood of coming n number of days earlier is 25 times higher than visiting n number of days later (Note that this skew factor is not learned for each patient nor from all patient visits). We use a predetermined value, but have observed that changing this value does not appear to have a significant effect on our Bayesian analysis. Given a medsRunOut value of 60 and alpha of 0.000039, Figure 7.3 shows the (unnormalised)



(a) Plot without skew factor.



(b) Plot with skew factor

Figure 7.3: Comparison the (unnormalised) likelihood of patient visits with and without skewing.

likelihood of patient visits without[4] and with[5] the skew factor.

As mentioned earlier, MCMC draws a large set of samples, and the sampling technique aims to draw samples proportionate to the probability distribution to be learned (the posterior distribution of (medsRunOut, alpha) in our case). Each element of the MCM chain is selected via taking a "step" from the previous element (or re-selecting the previous element). However, at the lower and upper ends of the medsRunOut interval ([0,89]) it is only possible to step in one direction. To prevent this from biasing the search, we updated our medsRunOut to a Uniform Distribution[0,(90∗n)-1] where n is a large number (we chose n=125 in my experiments). Then, we defined a variable medsRunOutmod which is the value of the sample of medsRunOut, modulo 90.

$$medsRunOutmod \sim Deterministic(medsRunOut \ \% \ 90)$$

The summary of our statistical model is :

$medsRunOut \sim DiscreteUniform(0, 89)$

$medsRunOutmod \sim Deterministic(medsRunOut \ \% \ 90)$

$alpha \sim HalfStudentT(\lambda = 0.04, \nu = 3)$

$patient\_visit \sim f(visit\_date) = e^{-alpha * distance\_between(visit\_date, medsRunOutmod)}$

[4]Figure 7.3a is plotted using the equation: $e^{(-0.000039 Min[(x-60)^2,(x-(60-90))^2,(x-(60+90))^2])}$

[5]Figure 7.3b is plotted using the equation:
$e^{(-0.000039 Min[(5^{Sign[x-60]}(x-60)^2,(5^{Sign[x-(60-90)]}(x-(60-90))^2,5^{Sign[x-(60+90)]}(x-(60+90))^2])}$

Our data has appointment dates since 2010-05-01, so we consider this as our Bayesian start date. The observed values are calculated based on this Bayesian start date as follows:

For each observed visit date for a patient, we define:

$$mod\_90\_visit = (visit\_date - bayesian\_start\_date) \ \% \ 90$$

We then run MCMC search to sample the posterior distributions given a sequence of mod_90_visit values as observations of the of the variable patient_visit. Figure 7.4 shows the trace of samples of the posterior distribution of the model parameters, given the observed values for the variable patient_visit (for a given patient). As mentioned earlier, each element of the MCM chain is selected via taking a "step" from the previous element (or re-selecting the previous element) as more data becomes available. Convergence of the MCMC trace is indicated by the traces having a "fuzzy caterpillar" shape.

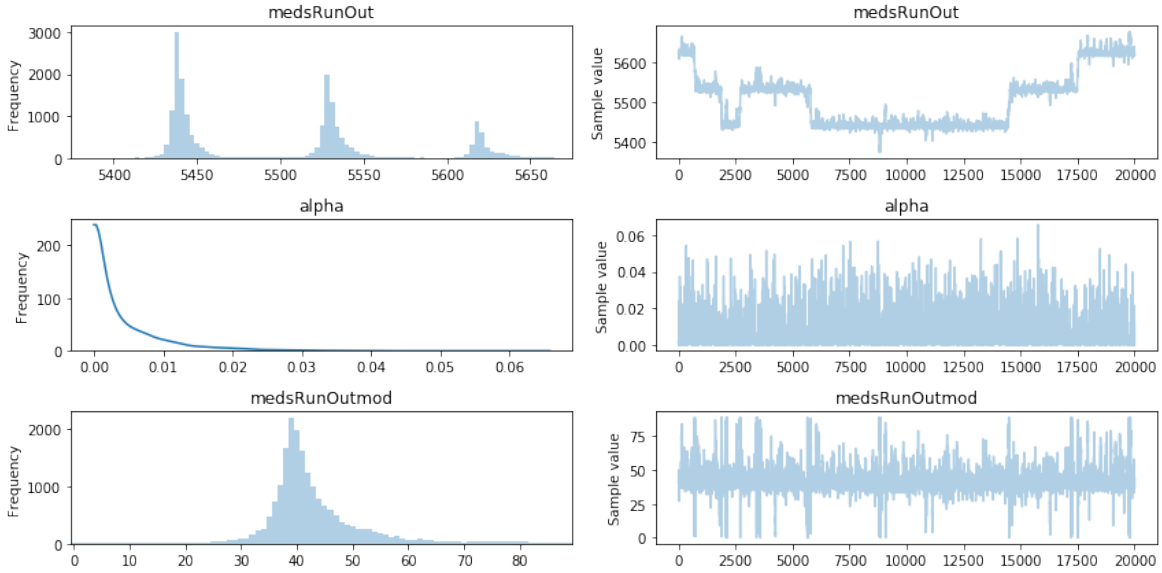


Figure 7.4: Trace plot for parameter posterior distribution samples given observed visit values for a patient (Run 1).

The following section will describe how this Bayesian inference was used in our prediction model. Note that each time the Bayesian inference is applied to the same set of observed values, there is some extent of uncertainty in the learned values. For instance, for the same patient, re-running the Bayesian model would give a different sample set as seen in Figure 7.5.

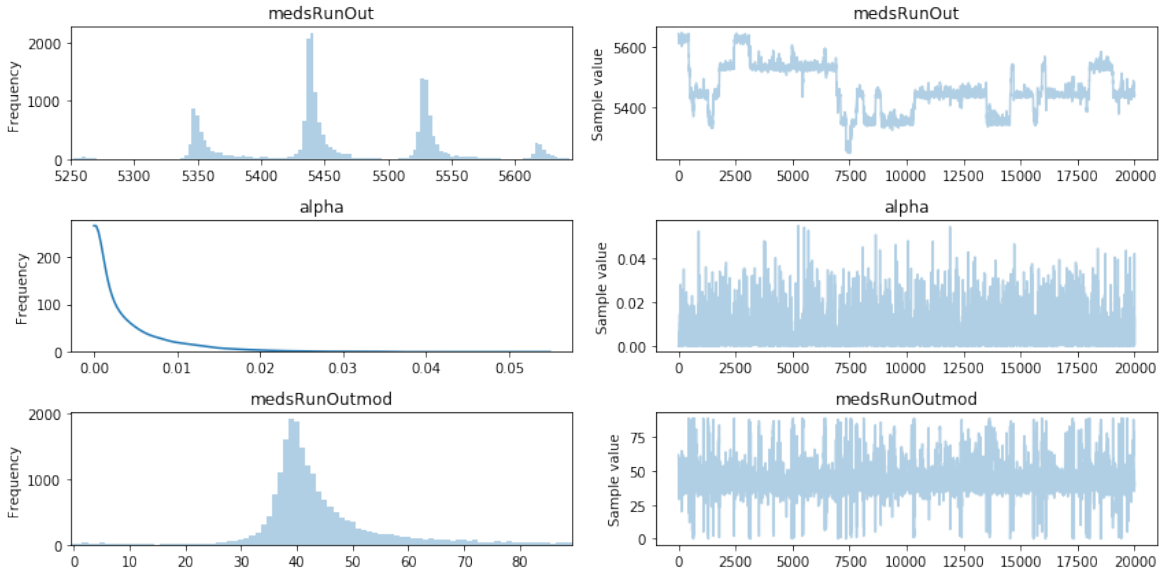


Figure 7.5: Trace plot for parameter posterior distribution samples given observed visit values for a patient (Run 2).

7.2 Prediction Model considering Bayesian inference

In order to apply the results of Bayesian inference in simulation of the patient visits, for each forecast recall for a patient we need to choose the expected visit date by sampling from the visit likelihood distribution (f) for that patient. We need to know which medsRunOut and alpha value to use for this sampling for each patient.

A pure Bayesian approach samples from the *posterior predictive distribution*:

$$P(x_{new} \mid \mathbf{x}) = \int_{\theta \in \Theta} P(x_{new} \mid \theta) P(\theta \mid \mathbf{x}) d\theta$$

where x_{new} is a new value, $\mathbf{x}$ is the set of observed values, and θ ranges over the set of possible parameter values Θ.

In our case, the model parameters are alpha and medsRunOut which we denote below as α and mro, respectively. Therefore,

$$\theta = (\alpha, mro), \text{ and}$$

$$P(x_{new} \mid \theta) = P(x_{new} \mid \alpha, mro) = \int_{\alpha, mro} e^{(-\alpha \ * \ distance\_between(x_{new}, mro))}$$

When using MCMC, because we are using a sample of the posterior distribution of model parameters, the integral above becomes a sum over sample values. In order

to use the posterior predictive distribution in our simulation, we would need to have the (large) sample available to the rule engine. However, having large samples for each patient (4 190 patients in our case) will take a large amount of memory in the rule engine and effect the efficiency of the model. Therefore, instead of using large samples for each patient, we use the common approach in machine learning of using a "plug-in" approximation to the posterior predictive distribution (Murphy, 2012):

$$P(x_{new} \mid \mathbf{x}) = P(x_{new} \mid \tilde{\theta})$$

where $\tilde{\theta}$ is a 'best' estimate for the model parameters from their posterior distribution $P(\theta \mid \mathbf{x})$ (or, when using MCMC, from the generated sample from that distribution). In this case, for each patient, we use a maximum a posteriori (MAP) estimate by choosing the mode of medsRunOut (denoted using $\tilde{mro}$) in the MCMC sample, and the central point of the 5% HPD (high probability density) interval for alpha values[6] in the sample that occur when medsRunOut = $\tilde{mro}$ (denoted using $\tilde{\alpha}$). The last equation above then implies using these parameter values in the exponential scaling function f (see patient_visit distribution in the summary of our model on page 139) as follows:

$$f(x_{new}) = e^{-\tilde{\alpha} \, * \, distance\_between(x_{new},\ \tilde{mro})}$$

Given a reference date, for each patient we learn the model parameters' posterior distribution and calculate $\tilde{mro}$ and $\tilde{\alpha}$. During simulation, we know that the next visit should correspond to next medsRunOut (= $\tilde{mro}$) date. Referring back to Figure 7.2, this means during visit v, the next recall will be corresponding to mro in Interval C (in the figure, denoted as mro+90 with respect to Interval B). We then normalise the f values across integer values in [0,89], and use those probabilities for sampling when a recall decision has to be made for a patient.

Before diving into details of the implementation of the Bayesian approach in our workload prediction model, let us look at the different reference dates referred to for Bayesian inference. For Bayesian inference, we chose a historic reference date as the Bayesian start date (01/05/2010 in Figure 7.6). The observed values for each patient visit date are then calculated as (visit date - Bayesian start date) modulo 90 (For example in Figure 7.6, a visit date 07/12/2011 is an observed value 45). The time line from the Bayesian start date is then viewed as series of 90-day intervals with a series of reference dates that each refer to the start of a 90-day interval. The most recent reference date that falls prior to our simulation start date is the initial reference date

[6]alpha is a continous distribution and therefore no mode can be calculated

considered to start the simulation of recalls. Thus, for Bayesian inference, the initial reference date, rd1, rd2, rd3, and rd4 in the figure are reference dates associated with patient visits to indicate the starts of 90-day intervals in which the visit should occur during simulation.

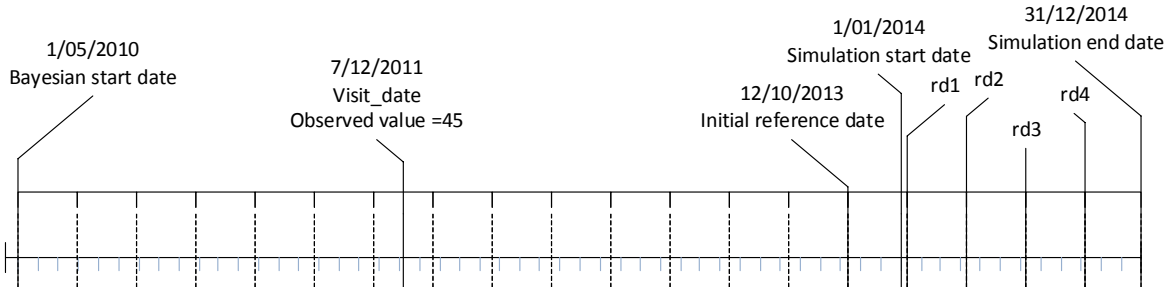


Figure 7.6: The time line reference dates in context for Bayesian inference.

We implemented the Bayesian approach in two steps:

Step 1 Pre-test using "cleaned" data of three-monthly recall patients.

Step 2 Applying the Bayesian model to all LTC patients.

7.2.1 A pre-test of using our Bayesian model in simulation

Initially we did a pre-test, as a trial to see if the Bayesian inference gives results according to our beliefs of how the patient visits happened historically. Since we assume that the patient visits are related to the 90-day medication prescription cycle, and the three-monthly recalled patients must be recalled each 90 days (following the care plan), we first test the approach by applying it to a subset (see below) of patients who are on a three-monthly recall period.

We first filtered data for LTC appointments of our cohort patients as follows:

1. Patients are on three-monthly recalls (i.e., the least recall period among various recall periods for each patient is 3 months).

2. Most consecutive visits for these patients fall around a 90-day cycle, from a chosen Bayesian start date; this Bayesian start date is the same for all patients. This ensures the observed data have least variance from the medsRunOutmod values for these patients, and thus we can check that the approach works in the simplest case where the data is clean.

3. When visits of a patient fall too close i.e., more than one visit in a 90-day interval, one visit is removed; this choice was made (manually) depending on how removing a visit date affects the overall visit pattern of that patient.

In other words, this makes learning easier for the Bayesian model to check if the approach works in the simplest case. This filtered data was then used in computation of the posterior distributions of parameters medsRunOutmod and alpha. Then the mode (i.e., the point estimate) of medsRunOutmod, and the central value of 5% HPD of alpha values corresponding to that medsRunOutmod mode value were calculated for each patient. Given these medsRunOutmod and the alpha values, the probability for each patient to visit on each day, represented as an offset from the reference date is calculated (refer to Appendix A.2 for complete implementation details). Thus, apart from their medsRunOutmod, each patient is also associated with pairs of each possible offset value and its corresponding probability for the 90-day interval.

When the day in our simulation happens to be a recall date for a patient, for each such patient:

- patient visits are generated,
- the reference date for that patient is advanced to refer to the next interval period;
- and, the next recall date is calculated as their (new) reference date plus the sampled offset value for that visit of the patient.

The start date of the 90-day interval (relative to the Bayesian start date) immediately prior to our simulation start date is chosen as the initial reference date for deciding when the patients are due to visit next (see Figure 7.6). There are chances that some of these visits may be prior to our simulation start date, in which case, the reference date for those patients is advanced to the next 90-day interval (referring to Figure 7.6 the reference date is updated from initial reference date to `rd1`) and the visit date is calculated from sampling again for such patients.

When selecting patients with three-monthly recalls, and with clean data matching our understanding of patient visits, the cohort reduced to 164 LTC patients. Figure 7.7 shows that Bayesian inference makes the workload prediction more realistic and reasonable because the periodicity due to three-monthly visits has disappeared and the line representing the predicted workload using the Bayesian approach is intermingling nicely with the historic visits of these patients.

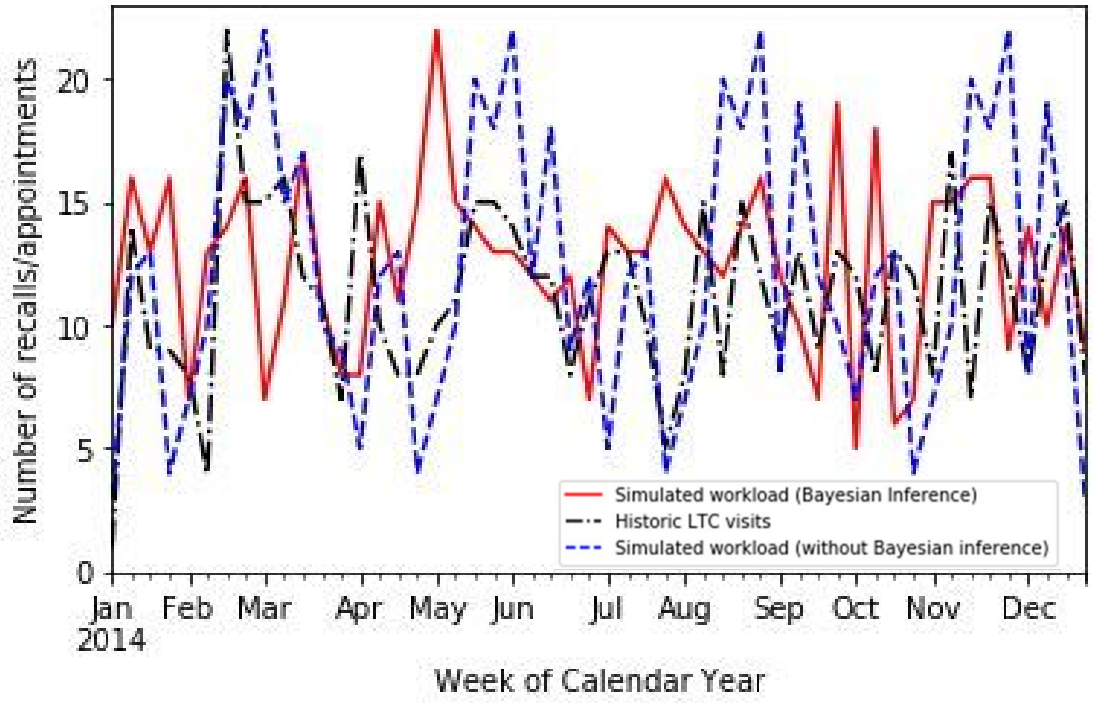


Figure 7.7: Simulation results for three-monthly recalled patients (clean data) [Run 1].

It should be noted that the differences between historic and predicted workloads are not solely to the method used to schedule recalls. Historic visits may also include visits not related to medications running out for a patient e.g., due to exacerbation of an LTC in a patient, and our model does not predict these "outliner" visits.

7.2.2 Applying Bayesian model to simulate LTC visits for all patients

Given that the Bayesian inference gave promising results for a subset of three-monthly recalled patients, we applied our model to learn the medsRunOutmod and alpha for each LTC patient in our cohort (i.e., for all 4190 patients). Similar to the three-monthly recalled cohort, we did a modulo with a multiple of their recall period, i.e., for three-monthly recall patients, we use modulo 90, and for six-monthly recall patients we use modulo 180. We also used reference dates that are Bayesian start date plus some multiple of 90-day intervals or some multiple of 180-day intervals for the three-monthly and six-monthly recalled patients, respectively. We then learned the best medsRunOutmod value and the alpha values for each patient as before (subject to these modifications). Finally, when a recall decision had to be made for a patient to

decide the next visit date, we use the distance_between function in Algorithm 3 with, instead of 90, the appropriate interval length. However, we still observed periodicity in the patient visit patterns (see the solid line in Figure 7.8).

Result: The distance v from mro in terms of 90-day interval
$dist = v - mro$;
$skew\_factor = 5$;
$value\_near\_mro = (skew\_factor^{sgn(dist)}) * dist^2$;
return $value\_near\_mro$;

Algorithm 3: The modified distance_between function.

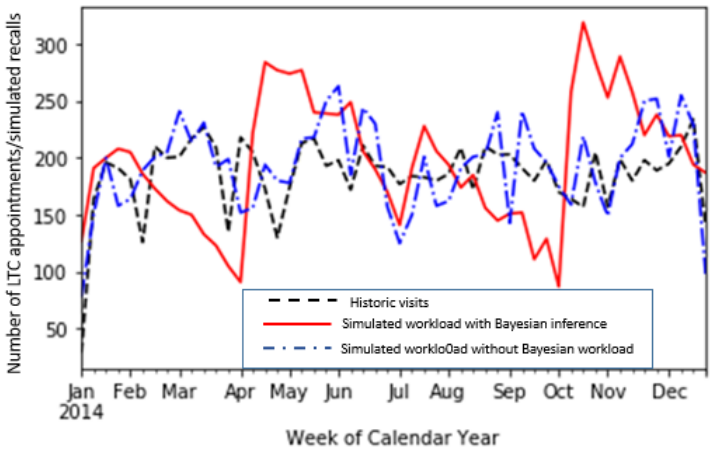


Figure 7.8: Bayesian inference applied to all LTC patients using their recall period.

There could be various reasons for this:

1. There were patients with more frequent recalls (the 1-monthly and 2-monthly) recalled patients. Their visits are more likely to be driven by their plan-of-care rather than their medications running out. Moreover, we have no information about how their medication prescriptions were issued or on which visit do they get their medication prescription. Therefore, in this simulation we recalled these patients based on their care plans.

2. Some historic visits might also be a reflection of a change in plan-of-care for some of these patients, which is not captured accurately in the data.

3. The modified distance function does not consider late or early visits.

First, we considered three cases based on a patient's recall period, to make a choice whether the most recent start date of the 90-day interval was referred to a reference date

corresponding to a visit for these patients or a due date for reissue of their prescription. For instance, in case of six-monthly recall patients, a reference date could be related to either an actual visit reference date (denoted as 0) or a medication renewal date (denoted as 1). This implies that, for six-monthly recalled patients, their visits and prescription reissue visits were alternative. Similarly, for 12-monthly recalled patients, there were four options for a given reference date: it could be a reference date for an actual visit (denoted as 0), the first medication renewal date (denoted as 1) the second medication renewal date (denoted as 2), or the third a medication renewal date (denoted as 3). This implies that after an actual visit, the patient gets the prescription reissued three times and the fourth time it is an actual visit to a GP.

Then, we used the distance function (given in Algorithm 2) that would predict, with respect to a recall reference date, a late or an early visit, which gave more reasonable results (see Figure 7.9).

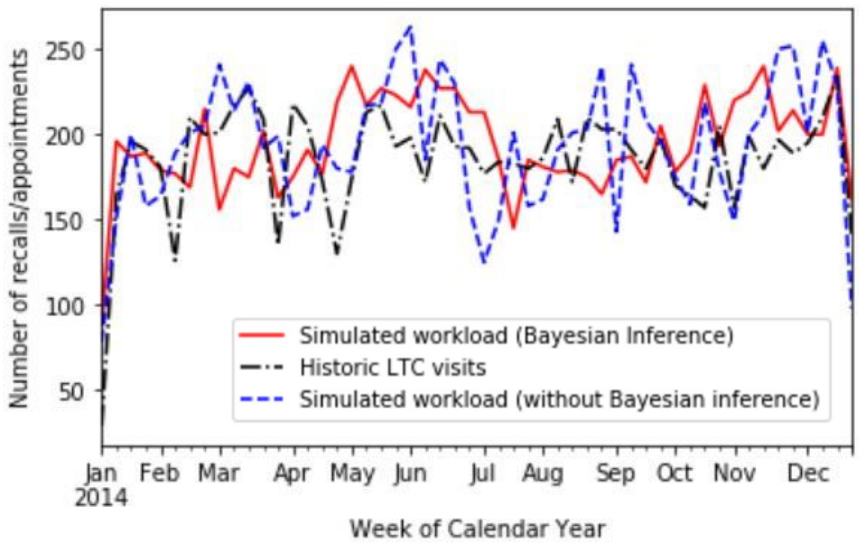


Figure 7.9: Bayesian inference (using distance function that predicts late or early visits) applied to all LTC patients using their recall period.

Given that medication prescriptions run out in 90-day cycles, we also conducted another simulation in which we applied the mod_90 (i.e., the number of days between the Bayesian start date and the date of visit) to encode observed values for each patient (we consider that the patient visit would have happened in some 90-day interval from the Bayesian start date). We seeded these observed values to our model and calculated medsRunOutmod and alpha values as described before. Given that our cohort included patients with 3, 6 or 12-monthly recall periods who required sampling to decide their

subsequent recalls, and we do not know if any reference date is related to a visit or a medication renewal date as described above, we approached this in two ways.

1. Method A: Given that six-monthly recalled patients have two options (0 and 1), and for 12-monthly recalled patients have four options (0,1,2, and 3), we randomly chose one option among these various options available depending on each patient's recall period. Figure 7.10 shows the simulation results using this method.

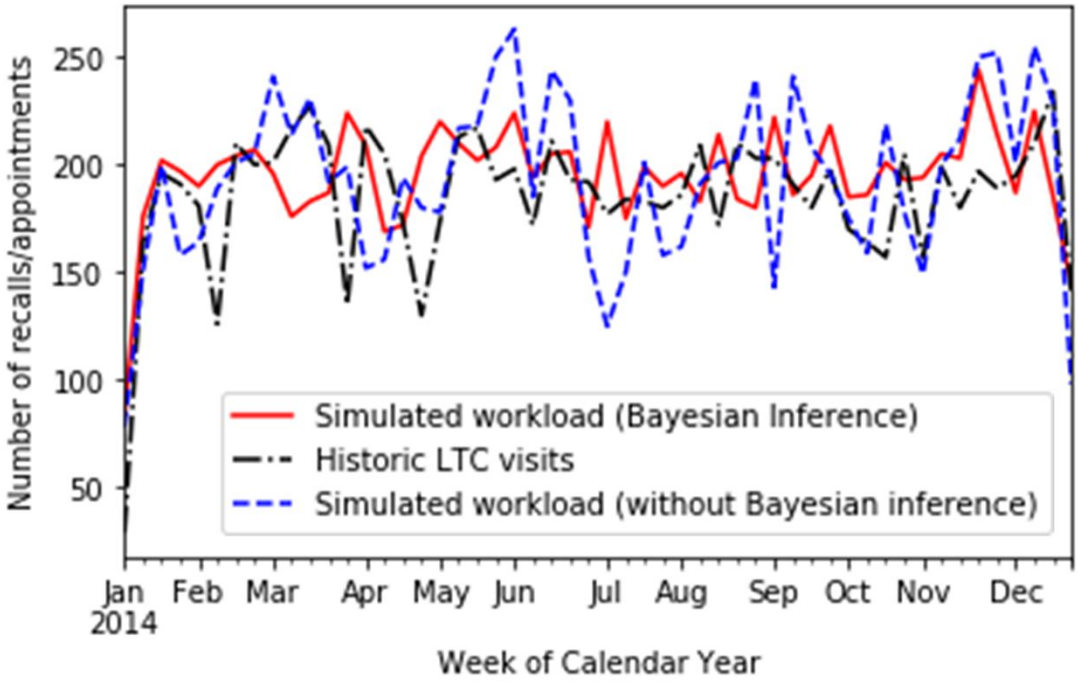


Figure 7.10: Simulation results for Method A.

2. Method B: We counted, from the Bayesian start date, how frequently a patient visit happens in the various 90-day intervals and chose the most frequent period as the visit period. However, there were cases where there was a tie between various 90-day intervals (especially for 12-monthly recalled patients), then we randomly chose a 90-day period from the most frequent periods for that patient. Figure 7.11 shows the simulation results using this method.

To this point, we emphasise that we use the Bayesian inference approach as a proof-of-concept to show that the workload prediction can be improved by learning from the patient's historic visits. Therefore, extending the approach to apply effectively to improving for all patients is out of scope for this thesis. Moreover, due to the inherent

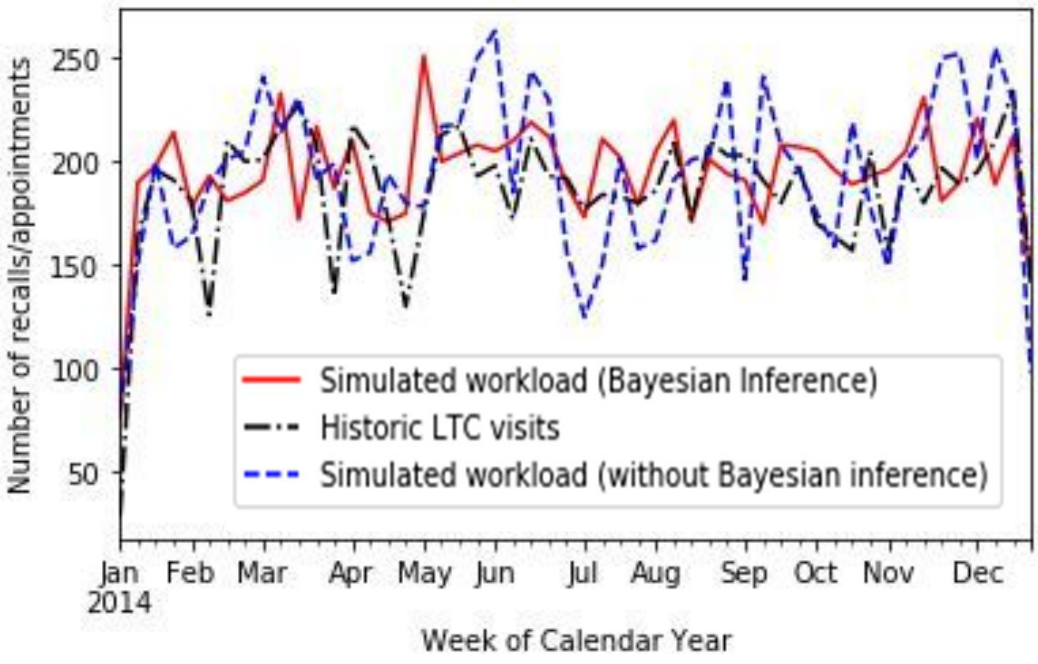


Figure 7.11: Simulation results for Method B.

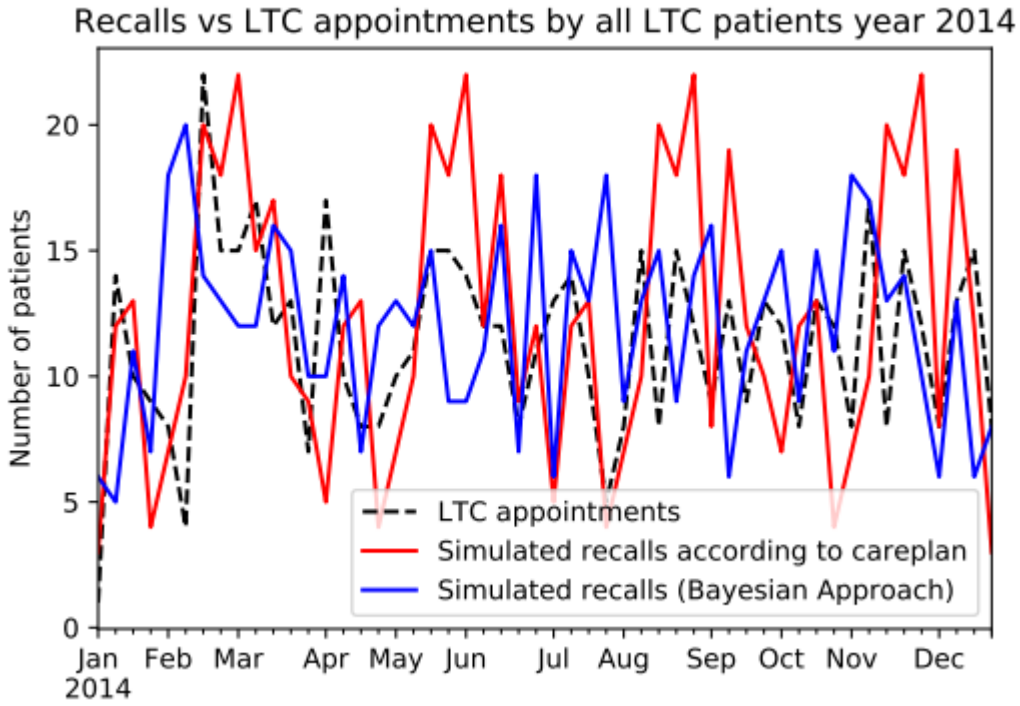


Figure 7.12: Simulation results for three-monthly recalled patients (clean data) [Run 2].

randomness of MCMC Bayesian inference and our use of sampling during simulation, the simulated workload varies each time the Bayesian model is applied to decide recalls (e.g., see Figure 7.7 and Figure 7.12). In our what-if scenarios, we want to compare simulation results between the workload generated by base simulation logic (our initial workload prediction logic) and the workload due to a change in practice (with our scenario logic added). Thus, we expect to analyse the impact of the scenario on the predicted workload. However, because each time Bayesian inference is applied in initial workload prediction logic, the initial workload itself varies across runs. Given, we want the differences in the workload that are due to the changed logic to be evident, in future, one could address this by allowing the rule-based system's working memory to be seeded with facts recording visit dates generated in a previous simulation, and adding rule logic ensuring that these previously generated dates are chosen for recalls, except where a scenario-based involves omitting or adding visits.

Chapter 8

Patient Information Model: Support for the population-level workload analysis

Revisiting Section 3.3, our research process has three phases, namely, information interpretation, rule-base development and improving simulation capability. As we have seen so far, these phases are intertwined. In Section 4.4.1, we presented the shortcomings of the dataset and the measures we adopted to compensate for these shortcomings. We had made some assumptions and definitions (through frequent and rigorous feedback cycles) in order to compensate for those shortcomings. This helped us identify what is required to support population-level workload analysis. We applied this patient information model in Cycle 4 and onwards of our DSR cycles presented in Chapter 3. Initial part of the chapter discusses current ontologies present in health domain. The rest of the chapter then presents our "three-cycle" DSRM approach (adopted from Hevner (2007)) and the developed patient information data model that supports workload analysis at a population level.

8.1 Ontologies in health domain

In the early days of advancement of IT in the health domain, from the perspective of a clinician, computer systems were developed to acquire patient data, and present it as a summary to the clinician to help him/her make an informed decision about the patient's health (Chute, 2000; Goertzel, 1969). Over the years, doctors developed their own terms and terminologies to help them store and share information and knowledge

among themselves (Chute, 2000). At present, the widely used clinical terminologies are Read codes (NHS Digital, 2017a) standardised in the UK (e.g., G20, C109), the SNOMED Reference Terminology (Spackman et al., 1997) developed in the US, the result of merging and expanding these two coding systems: the SNOMED Clinical Terms (SNOMED-CT) (Ministry of Health NZ, 2017j), the ICD-10 (WHO, 2016a) introduced by the World Health Organisation and the LOINC created and maintained by the US for laboratory codes. The International Classification of Primary Care (ICPC) (O'Halloran et al., 2004) is produced by the World Organization of National Colleges, Academies and Academic Associates of General Practitioners/Family Physicians (WONCA) now known as the World Organization of Family Doctors (Wonca). It classifies data as reasons for encounter (RFE), diagnosis or problem, and process of care. However, ICPC codes are more common in Australia than in New Zealand (Recently SNOMED-CT is more popular and Read codes were used prior to adopting SNOMED-CT).

All these clinical terminologies define a hierarchical classification system with codes associated with each entry (e.g., Figure 8.1 shows the SNOMED-CT design with its hierarchies). These codes are commonly used within electronic health records (EHRs) to provide a standard and unambiguous representation of a health condition, body part, medical test, etc. However, where and how this information is used remains ambiguous. We acknowledge that these clinical encodings are essential, however from a computer-assisted workload management's point of view, there is essential to understand where, why and what interventions were and are required for a patient.

Furthermore, shared decision-making requires storing, accessing and sharing patient information across various health IT systems which led to the development of standard health data models (Demski et al., 2016). While clinical terminologies were standardised for health care professionals, standard data models for electronic patient information were developed to allow data exchange between health IT systems (Tsiknakis et al., 2002). For example, the standards body Health Level Seven International (HL7) has developed a family of standards including the HL7 Reference Information Model (HL7, 2016, p. 7). This includes the key concepts of *entities* (e.g., people, organisations and places) acting in *roles*, while participating in *acts* (e.g., patient encounters, observations and procedures), which may have relationships with other acts.

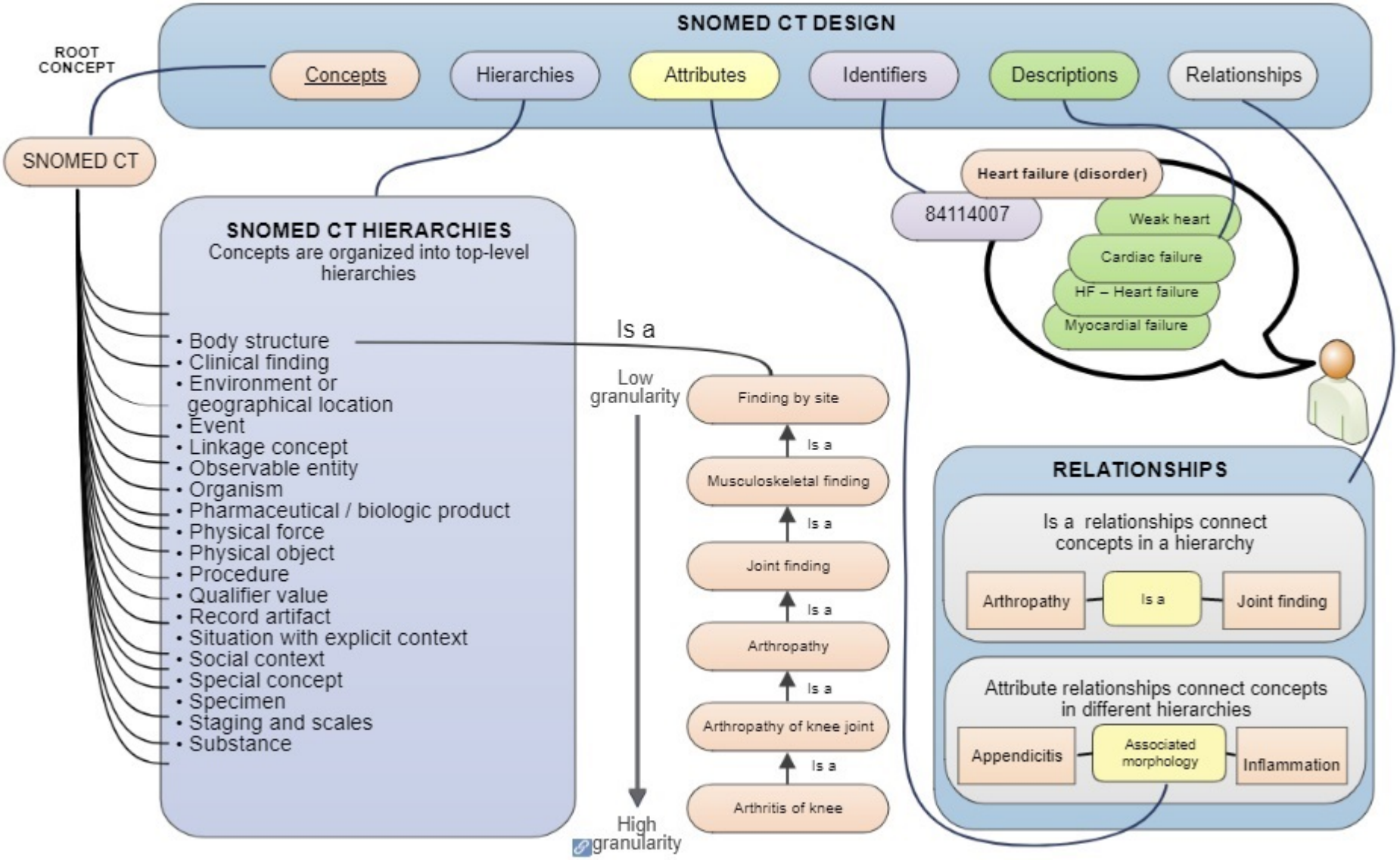


Figure 8.1: SNOMED CT Design from taken main page of SNOMED International (2018).

An entry in an openEHR Information Model provides a classification of entry types data types and data structures. However, the actual meaning of any term in the record cannot be known in isolation from the rest of the terms in the structure. This means neither openEHR IM nor its Data Model define relationships between the terms and the various levels in openEHR architecture. Furthermore, openEHR has an added functionality to classify plain text to coded terms. However, it does not capture or inform the clinician whether the same text with a mapping exists within the system. Both the HL7 and openEHR models can be specialised with domain-specific profiles (for HL7) or archetypes (for openEHR). However, such specialisations do not seem to exist for the area of primary health care.

The US Office of the National Coordinator for Health Information Technology has developed the Quality Data Model (ONC, 2016). This defines the structure for several types of data items related to the care of a patient, as well as a language for defining expressions expressing measures of the patients care, e.g., that certain medication is prescribed within a specified period after a given lab test. The models are intended to support clinical quality improvement processes. The data items modelled cover encounters (ordered and performed), diagnoses, medication that is currently prescribed, as well as medication dispensing and administering events, and laboratory tests, amongst others. The defined attributes include some relationships such as the diagnosis addressed during an encounter and the severity of a diagnosis or a symptom.

Mabotuwana and Warren (2010) proposed the ChronoMedIt framework to help practices to audit how well they manage their chronic patients. They found that identifying patients who receive suboptimal chronic care was difficult due to the temporal dependency of care actions for such patients. One of the challenges they faced was that the clinical encodings used in various practices varies, for example in primary health care sector, Read codes (NHS Digital, 2017a) were common in New Zealand while ICPC (O'Halloran et al., 2004) codes were common in Australia. Therefore, they used a *Problem* class to map these encodings to their generic names of chronic illnesses. For example, Read code C104.11 means a patient developed renal impairment due to diabetes and hence in ChronoMedIT, they mapped that Read code to Problem classes *Diabetes* and *Renal_Impairment*. We borrow this Problem class concept from this work.

In the following section we present our three-cycle approach to develop the patient information model presented in Section 8.4.2.

8.2 The three cycle approach

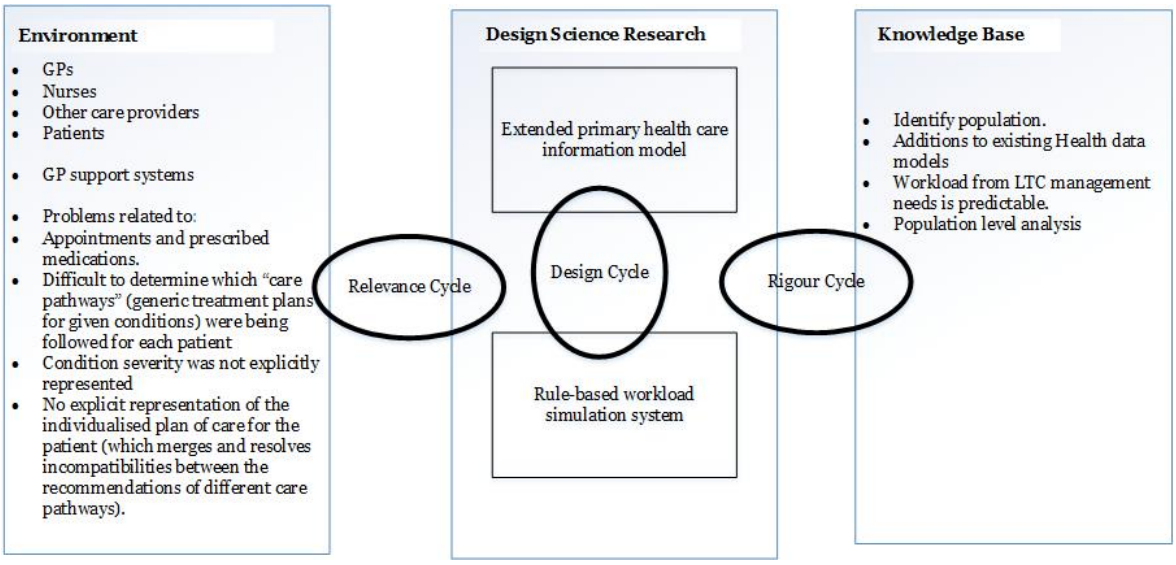


Figure 8.2: Hevner's three-cycle DSRM approach applied to develop our patient information data model.

Hevner (2007) introduced a DSRM approach with three-cycles, namely the Relevance cycle, the Design cycle and the Rigour cycle, and hence the name "three-cycle" approach. Figure 8.2 shows these three-cycles in our context. As depicted in Figure 8.2, the relevance cycle iterates to identify the environment and the problems in the context. The design cycle is the most important iterative cycle among the three-cycles. It iterates through the development of the prototypes or processes, evaluating them in the context. This cycle also considers the changes that may happen to the environment. In our context, the design cycle identifies various entities and the relationships between those entities that builds the patient information model to support population level analysis. These entities and the relationships help address the shortcomings identified in the environment. Section 8.4.2 discusses this patient information model in detail.

The third cycle, the rigour cycle, communicates the lessons learnt through the relevance cycle and the design cycle to the existing knowledge base. Through the rigour cycle, we identify the LTC population. Apart from making additions to existing health data models, we also added to the knowledge base that the workload from LTC management needs is predictable. We also added population-level analysis measures to the knowledge-base.

The rest of this chapter presents these three cycles and how their outcomes help us build our patient information model.

8.3 The Relevance cycle and the Environment

As depicted in Figure 8.2, the relevance cycle iterates to identify the environment and the problems in the context. Hence, this specifies various people, tasks, shortcomings and so on in the context. The major roles identified in this cycle were patients, GPs, nurses and other care provider roles at a PHC centre. Here, we focused on the shortcomings of data, mainly those related to the appointments and prescribed medications of their LTC patients. In relevance cycle we also discussed the inability of PHC data to capture the care pathways applied to the patients. Furthermore, the severity of conditions in a patient was not explicitly represented in the health data. There was also no explicit representation of individualised plan-of-care for the patient, which merges and resolves incompatibilities between the recommendations of different care pathways. This cycle iterates within the environment and the elements of DSRM (the extended primary health care information model and our rule based workload simulation system).

8.3.1 Primary Health (PH) data quality

Electronic Health Records (EHRs) or Personal Health Records (PHRs) capture the individual patient requisites to share the goals and activities set for the patient across various care providers without the loss of confidentiality (Burt et al., 2014; Mathers et al., 2011). However, these health records do not give an insight as to how to use that captured "digitized patient information" (Kohli and Tan, 2016) for a population-based reasoning. Moreover, according to Ministry of Health NZ (2017a), medical practices vary in

- compliance to the data capture methods in their clinical decision support systems (CDSS);
- GPs within practices vary in their mode of care delivery; although the data capture uses the same CDSS, the data captured regarding care delivered varies; and
- the established benchmark points; the performance matrices vary which means the data captured may be only in alliance to the benchmark metrics.

Hence, the data shortcomings discussed in this section may also be due to practice specific characteristics. However, the feedback from various roles at Pinnacle Midlands

Health Network[1] and the Clinical Director of Primary Care, Waitemata DHB, agree that most of these shortcomings are common across various practices under them.

8.3.2 The missing data puzzle for a population-focused care

As presented above, healthcare models such as HL7 and QDM, continue to focus on acute care. The structure of health systems reinforces the acute model with fixed 8, 10 and 15-minute appointment slots (Iacobucci, 2016). This underlines a reflex response to acute crisis interventions. A more systematic proactive response for the LTC population is required (Reilly, 2013; Wagner, 1995). However, primary healthcare information systems and their database architectures are structured to support acute 'siloed' unplanned care (Marshall et al., 2016; Salgado et al., 2016). In particular, while working with an anonymised database from a PHC practice, we have identified some information and relationships missing from that database that were required to facilitate population-level analysis, including the following:

- There was no information about whether a patient is an LTC patient, in a convenient format required for this study. We were required to identify the LTC patients in order to understand the LTC population demographics, and the upcoming LTC management needs of this population.

- There was no explicit representation of whether appointments and prescribed medications were related to acute conditions or to the management of LTCs in a patient. Because there was no explicit representation of LTC status associated to patients, our first work around was to identify those patients who book appointments for LTC related problems. However, there was no explicit information regarding this. Yet another workaround measure was to identify patients who are on LTC medications. However, information pertaining to which disease is targeted by a medication prescription for a patient was not captured in the patient data.

- It is difficult to determine which treatment path of a care pathway was followed for a patient. Given more than one care plan apply to multi-morbid patients, in order to understand the reason for a specific care action for a patient, it is essential to capture which path of care pathway was applied to a patient.

[1] An organisation that provides clinical decision support systems for practices under Waikato PHO.

- The severity of the LTC was not explicitly represented and may require an analysis of the results of lab tests over a period of time. It is known that when an LTC worsens in a patient, more frequent visits are required to manage the LTC in them. However, current patient data models do not capture the severity of a condition in a patient. It may be argued that the clinical encodings embed severity of a patient's health condition. But, from a population level reasoning point of view, what is required is the ability to explicitly capture the severity of LTCs in a patient. For instance, this would help to identify those patients who have severe diabetes and are at high risk of developing CKD in them.

- There was no explicit representation of the individualised plan-of-care for the patient, which merges and resolves incompatibilities between the recommendations of different care pathways. The ability to identify which LTCs or care plans apply to a patient is also not enough to reason at a population-level. We need to identify why and how a care action decision was made for a patient. This requires us to capture the plan-of-care of a patient.

Having highlighted the shortcomings, the next section will discuss how the workaround approach we applied to overcome these shortcomings helped us develop our patient information model, aligned with the development of our workload prediction model.

8.4 The Design cycle

The definitions and assumptions unique to this work are a consequence of the inference logic applied in our context[2]. However, we also identify that the data needs to have the capability to capture patient information that can be used for a population-level workload analysis.

8.4.1 The information inference logic

To this end, I have introduced the shortcomings in the dataset that hindered us in developing a workload prediction model without assumptions. We had to infer knowledge to fill the gaps due to the missing information in the dataset. For instance, in order to find how many diabetic patients are treated with more than two LTC medications, we had to find which patients are classified as diabetic and the LTC medications they are

[2]Following design science research methodology, this inference logic follows the guidance from our PCA.

on. In this section, I discuss the inference logic and how it helped design our patient information model. I also would like to highlight that this patient information model has been used in our workload prediction model.

The first challenge was to identify LTC patients[3]. Although the information captured was not in a convenient format for this study, every patient was classified using Read codes. Hence, we mapped the Read codes to problem class(es) as described in Chapter 4 and tag them as LTC patients. Therefore, in addition to the clinical encodings, the health data should also capture the generic name (problem class) of the disease.

In order to simulate patient visits, we need to seed their visit details that instigate the subsequent patient visits in the simulation. However, such information related to LTC appointments was also not explicit in the data. Therefore, the next step was to identify the LTC appointments. The cause of appointments was not captured in the patient data. Therefore, we had to identify LTC appointments based on the recommendations (guidance was from experience as a GP) of our PCA.

According to our PCA, in order to address LTC management needs, an LTC patient usually makes appointment with a GP or a nurse. It that they make an appointment in order to get their medication reissued. Therefore, as presented in Section 4.4.1, our definition of LTC visits depended on active medications. We acknowledge that, in our case, the definitions and assumptions were driven from the perspective of a general practitioner and it may vary from person to person and from practice to practice. Therefore, we add a validity period to every prescription, and also to capture an LTC status that would denote whether the appointment was related to an LTC or not.

In the current GP support system (which aids the PHC practice whose data we use in this work), if, during a consultation, medications to manage LTCs in a patient are prescribed, then a MedTech32 'T' status is tagged with all the medications prescribed during that consultation for the patient. However, the data does not explicitly capture the condition for which the medication is prescribed, nor the validity period of the prescription. It is important to know the LTC status or the condition addressed by that medication and the validity of medications because, generally, an LTC prescription may be valid for up to four months, and there are medications such as paracetamol, that could be prescribed for both acute and long-term conditions.

We also assume that as an LTC worsens, the recalls for a patient are more frequent. Although this severity might be embedded within the clinical encodings, for

[3]The main focus of this work are the LTC patients enrolled at the primary healthcare centre.

a population-level understanding of the LTC patients, we need to capture severity in terms of low, medium, high and very high risks. We demonstrated with a what-if scenario in Section 6.1.2 that having severity explicitly represented can give a tool the flexibility and adaptability to shift a group of patients to other care provider roles. In another context, having an understanding of severity of conditions could help stratify patients based on their needs and plan care accordingly (Pines et al., 2015)

Every LTC patient may be on one or more care plans depending on the multi-morbidities in a patient, and so may refer to more than one recall period. However, every patient should have only one plan-of-care, which considers all the LTCs present in the patient. Every patient, thus, should have only one recall period at a time.

Furthermore, the information required to draw conclusion on a plan-of-care for a patient is not addressed in the existing health data models. As presented in Section 5.2.2, in our workload prediction model, during a patient visit the care pathways unfolds into a patient-specific plan-of-care. This step highlights the need to capture the care plans applied to a patient so that the clinical decisions can reason about care actions on a population-level. The data should also capture information to identify the criteria for a population cohort and thus can support population-level analysis.

The next section presents our patient information model.

8.4.2 The Patient Information Model

In this section, I introduce our ontology, and the basic concepts and relationships of our ontology. We borrow a few basic concepts and relationships from openEHR (2016), HL7 RIM (HL7, 2016), SNOMED-CT (SNOMED International, 2018), and a paper by Mabotuwana and Warren (2010). Figure 8.3 represents those concepts and relationships needed for a population-level workload analysis as an ontology modelled in OWL (OWL Working Group, 2012). The additions to the existing patient information model are highlighted using bold rectangles around the class. The green rectangles are instances, and the dashed lines denote the 'instance-of' relationship. The hollow arrow-headed lines shows the 'is-a' relationship between the classes.

The `isLTC` attribute shows the LTC status of the entity. Box A shows those entities relating to LTC (e.g., Appointment, Classification). The `patientId` is an attribute of a patient, plan-of-care, appointments, classifications, care actions and population. Box B shows this relationship. Boxes A and B thus explicitly indicate that we add this attribute to existing classes.

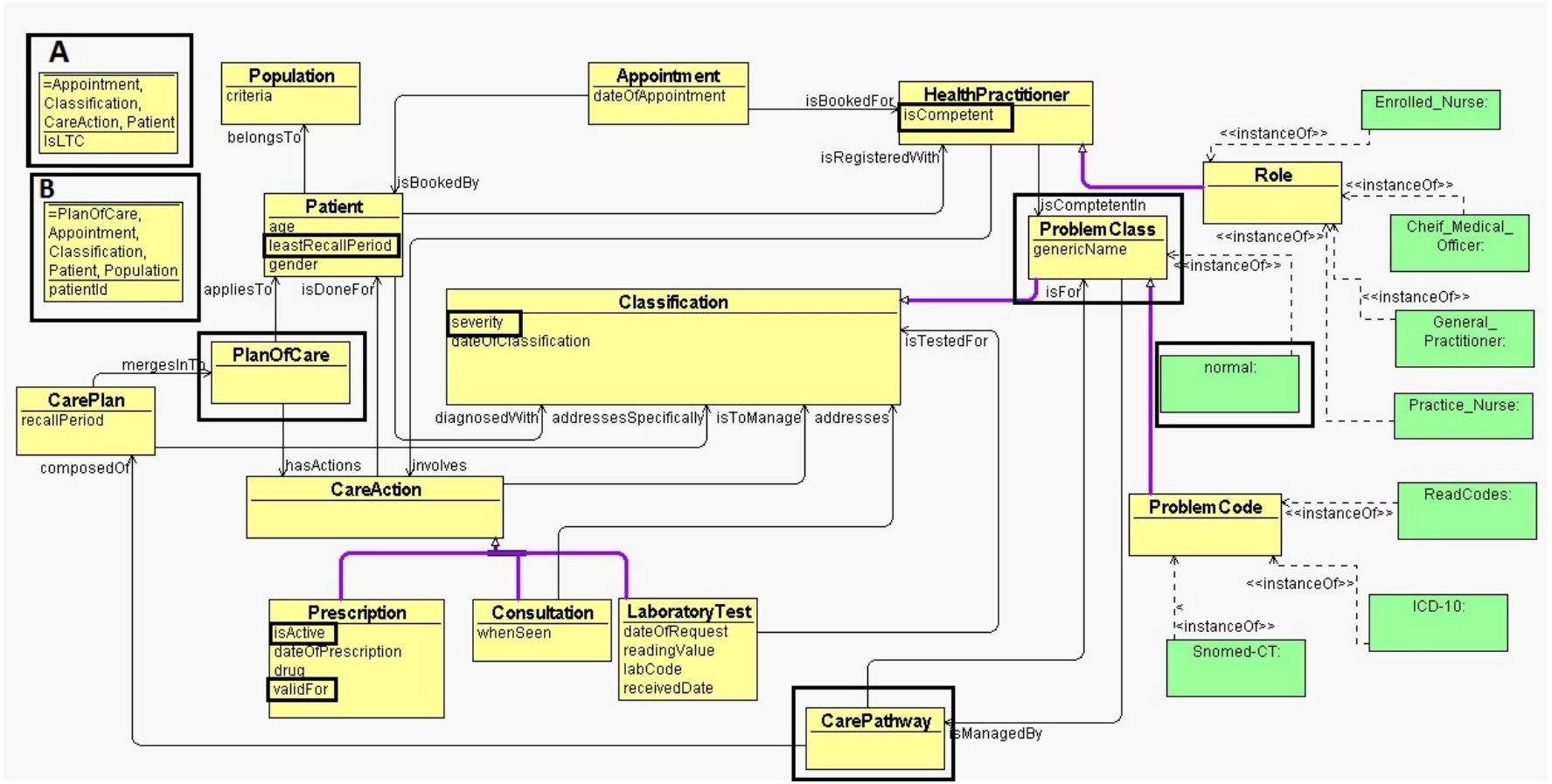


Figure 8.3: The Patient Information Model.

One major requirement is to identify the population. Depending on the criteria, the population varies. Therefore, we tag a patient, based on some (defined) `criteria`, `belongsTo` a population.

A health practitioner may hold various `role`s within the practice. These roles, such as general practitioner, practice nurse and chief medical officer, in a PHC context are shown as instances of a role. A health practitioner can be competent to manage a specific health condition, denoted by `isCompetentIn` a `ProblemClass`.

A patient `isRegisteredWith` a health practitioner and may have an encounter with a general practice via an appointment. Hence, an appointment `isBookedBy` a patient and `isBookedFor` a health practitioner. With these two relationships, we provide the flexibility for a patient to be registered with a health practitioner and can make an appointment with another health practitioner if required. During a consultation, a patient may be `diagnosedWith` a classification or addresses a classification within a patient. Every classification of a patient will have a diagnosis with a `severity` and the date of classification, which help to mine the medical history of a patient. A classification for a patient will thus, have the `patientId`, `isLTC` status, `ProblemClass`, a `severity` of the condition, and the `ProblemCode` of the condition along with the `dateOfClassification`.

A `ProblemClass` will have `ProblemCodes` associated with it. These `ProblemCodes` can be Read codes, SNOMED-CT or ICD-10 codes. We also emphasise that `ProblemClass` is not a replacement for clinical codes like Read codes, but is additional information and hence it provides more generic information required by both clinical (GPs and nurses) and non-clinical (clinical decision support developers) users of the clinical data. Furthermore, in our ontology, if a patient does not have a medical condition, the patient is classified as a normal patient. So `normal` is shown as an instance of `ProblemClass`. This enables the health system to identify normal patients who follow the generic care pathway for prevention and early detection of LTCs.

One of our major contributions through this ontology is to enable to identify the care plan applied to a patient. To determine the enacted care plans in a patient, we broadly classify the patients using `ProblemClass`. A `ProblemClass` (which has a genericName) addresses a medical condition like diabetes, and thus refers to various care plans for its care pathway. A care pathway specifies the guidelines to manage a specific LTC in a patient. In our ontology we define a relationship that captures this association between an LTC and its care pathway. Therefore, a ProblemClass `isManagedBy` a care pathway and a care pathway `isFor` a problem class.

We analysed the care pathways, derived the care plans based on the interventions required and the period within which the intervention must be made in a patient, captured as a `recallPeriod` within a care plan. For example, a newly diagnosed patient with a classification C109 (a Read code that classifies as Diabetes) with a reading value greater than 55 for 44TB (Read code for HBA1C blood test), may be monitored on 3-monthly basis while a patient whose diabetes is stable with the same HBA1C and who is on medication, will be monitored every 6 months. This intervention based on `severity` is captured as different care plans. Thus, a care pathway is `composedOf` care plans. Each care plan `addressesSpecifically` a classification.

The `ProblemClass` and the `severity` of the condition in a patient identifies the care plan for the patient. Every LTC patient may be on one or more care plans depending on the multi-morbidities in a patient, and so may refer to more than one `recallPeriod`. However, every patient will have only one plan-of-care, which considers all the LTCs present in the patient. Every patient, thus, will have only one recall period, the `leastRecallPeriod`. The most severe condition will require interventions that are more frequent and hence the plan-of-care considers that the most severe condition drives the recalls (Amir et al., 2015; Burt et al., 2014; Dennis et al., 2008). However, defining a `leastRecallPeriod` enables a clinician to use his expertise (i.e., the evidence-based care) to recall a patient as required than recommended by the plan-of-care for the patient.

The plan-of-care `hasActions` from care actions. A care action is a medical activity that `isToManage` a health condition in a patient. A care action `involves` a health practitioner; for instance, a GP prescribes the medication, and a nurse does lab tests. Therefore prescription, consultation and laboratory tests are care actions.

We have also considered to represent if a prescription for a drug `isActive` and `validFor` a certain period. For instance, a prescription for paracetamol may be active for a week i.e., the patient is on medication for a week but the validity of the prescription could be for 3 days, i.e., the patient will require a new prescription, if they need to collect the medication after three days from the date of prescription.

A consultation could address an acute condition or, one or more LTCs in a patient. In order to capture this reason for an encounter of a patient with the PHC, we define that a consultation `addresses` a classification.

Laboratory tests may be requested in order to rule out a diagnosis of a suspected health problem or to confirm the health condition of a patient. Therefore, we define that a laboratory test `isTestedFor` a classification.

It is important to capture which care actions are done for a patient. In order to capture this information, we explicitly capture that a care action `isDoneFor` a patient, and a plan-of-care `appliesTo` a patient. Aggregation of various care plans into a patient-specific plan-of-care ensures that medical activities, such as medications and lab tests, are not duplicated for a patient unnecessarily (Burt et al., 2014). Thus, during the consultation, a health practitioner considers all the medical conditions present in the patient to prescribe medication or order the lab tests. At times, this would also enable a patient to book the follow-up appointment. In other words, the various care plans `mergesInTo` a plan-of-care and the care actions depend on this plan-of-care for a patient. We highlight that this information required to draw the conclusion on a plan-of-care for a patient is not addressed in the existing health data models.

8.5 The Rigour cycle

We borrowed some of concepts from existing ontologies. Table 8.1 presents a summary of concepts of our patient information model borrowed from other work. Through the rigour cycle, we identify the LTC population. Apart from making additions to existing health data models, we also added to the knowledge base that workload from LTC management needs is predictable. We also added population-level analysis measures to the knowledge-base. Though we discussed this ontology from the perspective of LTCs, this ontology can support acute or non-LTC specifications too. Care actions

Table 8.1: Summary of concepts of our patient information model borrowed from other related work.

Our concept	**Concept presented as in other work**
Role of a health practitioner	role (from openEHR (2016))
Care action	act (from openEHR (2016)), care activity (from HL7 (2016))
Severity	severe (from Mabotuwana and Warren (2010)), severity (ONC, 2016)
ProblemClass	problem class (from Mabotuwana and Warren (2010)), openEHR (2016))
ProblemCodes	problem codes (from Mabotuwana and Warren (2010) and HL7 (2016))

could relate to non-LTC events like screening for pre-diabetes, annual health check-up or even encounters due to accidents. We emphasise that these additions to the existing ontology are highlighted with rectangular boxes in Figure 8.3.

In summary, following a three cycle approach adopted from Hevner (2007), we identified shortcomings in the current health data models and the data shared with us. In the process of compensating for these shortcomings through information interpretation, we articulated what is needed in a patient information model in order to support population-level workload analysis. We also discussed how this patient information model contributed to the knowledge base. We emphasise that this patient information model was an outcome of designing what is required to develop a workload prediction model for population level workload analysis. This patient information has been applied and evaluated through our workload prediction model.

Chapter 9

Conclusion

This thesis has addressed the challenge due to growing demand for LTC care at a primary health care (PHC) centre. Our adaptable best practice based workload prediction model (ABP-WPM) predicts the upcoming workload as a volume of appointments due to LTCs at a medical practice. This thesis also demonstrates the ability to analyse the impact of resource bound policies, external factors, changes in practice policies and adopting new models of care on the predicted workload, and evaluate the strategies to address them. The patient information model presented in this thesis addresses the shortcomings of current health data models to support a population-level workload analysis. This thesis also shows that with a better understanding (using Bayesian inference) of a patient's prescription renewal cycles and their likelihood of visiting early or late for an appointment improves the workload prediction capability of the model.

9.1 Learnings from related work

Based on the literature, for example work by Babiker et al. (2014); Harris and Zwar (2007); O'Leary et al. (2013); Wagner (1995), we realised that understanding and managing workload in a health care domain is a known challenge for decades. There have been attempts to understand the workload based on historic data (Abdel-Aal and Mangoud, 1998; Green Cross Health, 2017; Murray and Berwick, 2003; Potts et al., 2011). We focus on LTC management in a population of patients. The chronic care model (CCM) by Wagner (1998) serves as a guide to adopt measures to improve LTC care for patient population at a PHC level. We also learnt that adherence to clinical guidelines and a care planning process to address multi-morbidity needs of a patient are essentials for effective LTC care at a PHC level. However, few studies have focused

on adherence to guidelines, considering all LTCs at the level of an entire PHC practice. Table 9.1 gives a comparison of this related work to our work (refer to Chapter 2 for a fuller discussion on these work).

Table 9.1: Features of related work that motivated our approach to built the workload prediction model, and how they differ from our work.

Related Work	**Features that motivated our work**	**Contrast of the related work with ours.**
Anonymous (2006b); Babiker et al. (2014); Bodenheimer et al. (2002a); Harris and Zwar (2007); O'Leary et al. (2013); Reeves et al. (2014); Terry (2017); Trindade and Pires de (2013); Wagner (1998)	Similar motivation, need for care planning at a PHC level	Some studies considers specific patient population to analyse care planning at a PHC level.
Hall (2012)	Addresses scheduling resources; Study attempts to answer questions very much similar to our research questions.	Has more of a patient perspective. Questions in their study focus on the emergency department of hospital. Ours is a PHC context.
The CREDO framework (Fox et al., 2006)	Use of care plans to refer to personalised schedules and follow-ups of patients.	The CREDO framework focuses on how the services are delivered rather than anticipating what services will be required. We focus on anticipating workload.

Mathers et al. (2011)	Collects and aggregates data to determine unmet health care needs of the population	For them, the main reason for aggregation is to integrate self-management into their care planning process. We aim to help the practice equip themselves according to the predicted workload.
Brown et al. (2018)	Addresses how care planning is applied in a PHC context	Their work has a patient perspective while our work has a care provider perspective.
Abdel-Aal and Mangoud (1998); Green (2013); Green and Savin (2008); Murray and Berwick (2003); Potts et al. (2011); Utley and Worthington (2012)	Attempts to forecast demand for care for a population of patients	Uses mathematical or statistical models and has little or no discussion about evidence-based care or patient centredness.
Grant and Greene (2012); Pines et al. (2015); Reid et al. (2009); Struijs et al. (2015); Taylor (2015); WHO (2016b)	Analyses how services are provided.	Some studies focus on population-based care, while others focus on patient-centred care. Seldom do they address both, or otherwise they address only a selected few chronic conditions. In our work, we address how multiple care pathways can be integrated to predict population-level workload.
Davy et al. (2015); Harris and Zwar (2007); Trindade and Pires de (2013); Walley et al. (2008)	Analysis of impact of changes on workload	Presents the findings from a patient perspective. Ours is a population-level perspective.

Hawkins and Novak (2011)	Presents benefits in care from shifting patients to a trained nurse	Considers only diabetic patients; it does not mention increased or decreased workload due to the changes. In our work, we address the workload from following best-practice for LTC management (not only diabetes), and also explore various ways to manage this predicted workload.
Anooj (2012); Azadmanjir et al. (2017); Davis et al. (1977); Kahn et al. (1991); Lhotska et al. (2001); Papageorgiou (2011); Pawlak (1997); Pestotnik et al. (1996); Salatino et al. (2016); Shortliffe (1974); Shortliffe et al. (1975)	All use rule-based systems for clinical decisions. Some studies involve expert opinion and refer to guidelines to develop their rule base.	The rule representation varies. Some studies are specific to hospital settings. Ours is a PHC context. We also involve expert opinion to develop best practice guideline rules to predict the workload.
The WEB planning model (Segal and Leach, 2011)	Stratifies population; highlights problems with quality of data; involves clinicians and other care provider roles in developing best-practice for their population	Focuses only diabetic patients; suggests a unique clinical team for each patient and then aggregates to estimate total work force required. In our work, we consider multi-morbidity in patients. Instead for suggesting a clinical team, we aggregate workload to provide the practice the autonomy for their workforce development.

WiPP (Anonymous, 2006a)	Workload analysis tool	More details not available; not a generalisable tool; The ability to explore what-if scenarios makes our model an adaptable, generalisable one.
The Holon Framework (Warwick and Bell, 2007)	Uses what-if scenarios in a health care planning context. They view health care planning similar to a software process that can be evaluated using Goal-Question-Metric (GQM) method. The framework considers a transition from one state of patient health to another through interventions that address the who, what, where, when and how elements of care planning.	The impacts of such macro-interventions are examined using what-if scenarios. In our case, we use what-if scenarios to predict the impact on the predicted workload due to changes in practice, adopting a new model of care, posing a resource bound policy for attending LTC patients, and external factors. We do not consider the impacts at a patient level.

The iCare project (Indian Health Service, 2014)	Workload management issue addressed	Considers only specific events such as cervical screening and prostrate screening; a patient chooses if they require to be alerted about future health care events applicable to them. Or work aims to help the practices to foresee the future appointments for LTC patients, rather than alerting patients about their appointments.
Maher et al. (2009)	Refers to the need for a standard protocol to deliver structured, good quality care.	They do not address how these requirements can be met by the organisations. On the other hand, our tool refer to the best practice guidelines and can address such requirements; knowing the population and volume of sub-population visiting, the practice can plan and organise care actions accordingly.

Busetto et al. (2017); Massimi et al. (2017)	They explore various contexts and the impacts of involving nurses in the care planning and delivery process at a practice level.	Explicitly mentions roles and responsibility that the nurses can perform. We do not explicitly assign responsibilities, however, we examine the impact of shifting patients to nurses on the predicted workload. In our work, we assume they can attend the patients shifted to them.

A PHC system needs a system that can consider best practice guidelines for its patients registered with the practice. Best practice guidelines are clinical guidelines that are varied to fit the patients' care needs (Johnson, 1997). Therefore, adherence to best practice guidelines for the LTC patients at a PHC level will help address evidence-based, as well as patient centred LTC care with a population-level perspective.

9.2 Addressing what is required in a data model.

In the process of building our initial patient model (the individual patient data and a visit date to drive the simulation of patient recalls for a patient), we identified shortcomings in the dataset supplied to us, as well as in the current health data models, to support population level workload analysis. We followed a rigorous build, self-evaluate and feedback (from our PCA) cyclic approach to make definitions and assumptions to compensate for the problems due to missing data. This information inference phase helped us develop care plans (e.g, CKD care plans) as rules, identify patient information (e.g., the LTC status of patients and their appointments) and identify a visit of a patient to use to simulate the subsequent recalls for the patient. This information inference phase also helped us to identify what is required in a patient data model to support population level workload analysis. Thus we developed the patient information model and applied it in our workload prediction model, presented in Chapter 8.

9.3 The choice to use a rule-based system.

As the data related to which care pathways were applied to LTC patients was missing in our dataset, in our ABP-WPM, we have best practice guidelines in the form of an encounter-based unfolding plan-of-care. This encounter-based unfolding plan-of-care has layers of rules. The first layer is the disease specific care pathways (best practice guidelines for the practice, data for which is used in this study) which unfolds to disease specific care plans. These disease specific care plans (as rules) when seeded with patient data unfolds to form individualised plans-of-care. This process of unfolding of care pathways to individualised plans-of-care occur during the (simulated) patient visit. Therefore, although there are alternative approaches such as agent-based models and discrete event simulation to model health care roles and activities (Luke et al., 2018; Mustapha et al., 2018), a rule-based approach seemed the most appropriate in our context.

One of the strengths of our model is also that it is rule-based. Minutolo et al. (2017), Shiffman (1997) and Shortliffe (1974) lists (along with many others) modularity, openness, flexibility and the ability to unwind based on data set as strengths of a rule base system. In Chapter 3, we showed how following the design science research methodology (DSRM), each phase of development of our workload prediction model extended our rule base. We also demonstrated (in Chapter 6) how extending and adopting that rule-base can improve the capability of our workload prediction model to support various what-if scenarios. These features can help a PHC practice to know the upcoming workload, and understand the impacts of various strategies they may adopt on them as an organisation.

9.4 The holistic view of our approach

Figure 3.3 presented the three phases, namely II, RBD, and SC, and the details of their interactions required to develop our ABP-WPM. Although, in this thesis (except in Section 3.2.1), we presented these three phases as standalone interactions, these interactions were intertwined. For example, we needed to identify LTC patients. However, the data lacked this information, and so initially, we defined a patient as an LTC patient if the patient has made a GP or a nurse appointment. But the simulated workload was bizarre with periodicity, sharp peaks and dips. Then, analysing the simulated workload, our PCA, we considered the medication status of the patients. The data

lacked explicit representation of which LTC or disease is targeted by that medication. So, we defined active medications and then defined LTC patients based on active medications, which improved the simulated workload. The rule base development (RBD), thus had interactions with information interpretation (II) phase. Improving simulation capability (SC phase) thus had interactions with both RBD and II phase.

The following section highlights the contributions of this thesis.

9.5 Contributions of this thesis

In this work, we were informed by both the literature (for the requirement for a system to predict the workload at the medical practice) and by our PCA who is the collaborator in this work. The collaboration was in the form of data and the knowledge base to understand primary health care domain, and practical issues at the medical centre. Therefore, we followed the design science research (DSR) methodology. In Chapter 3, we presented how each phase of development of this work is driven by feedback from our collaborator (the PCA), primary health organisations (PHOs, specifically the Waikato, and the Otago and Southern Island PHOs) of New Zealand, members at Mosgiel PHC centre, and review comments on our published papers, to develop our workload prediction model.

9.5.1 Answering our research questions

In this work, we address how a medical practice can become proactive in delivering chronic care through population-level workload management through three main research questions as discussed below.

RQ1 Given a medical practice follows best practice, what model(s) can be used to predict the population-based care workload?

This question specifically focuses on the workload that is expected to arise from the demand of care to meet the LTC management needs in a PHC context. We demonstrated that using encounter-based unfolding plans-of-care for LTC patients can predict the population-based workload.

RQ2 How can this predicted workload be analysed in various what-if scenarios? For example, if the low severe LTC patients are attended by a nurse, what is the change in workload of the GPs?

We showed how our rule-based workload prediction model allows simple adaptation to provide the capability of analysing what-if scenarios. We first showed that assigning a certain number of slots per weekday per available GP can reduce the gap between the predicted workload and the availability of GPs to meet the LTC care demand. We showed two scenarios where patients are shifted to nurses. Through these scenarios we demonstrated that shifting patients from the GP workforce to a nursing workforce makes GPs more available for other patients. We also showed, using a simple financial analysis, that shifting patients to nurses would also bring more returns to the practice.

RQ3 How can the impact of various health policies be studied at an organisational level? For instance, how would the policy to have annual CKD screening for LTC patients affect this practice?

We considered the scenario of adding new care plans for high need LTC patients, where they are recalled at least in every three months. We also considered in another what-if scenario where patients are required to have a CVD screening test annually. Yet another scenario considered increasing GP and nurse slots for LTC patients by offering LTC appointments on Saturdays. Other than the impact on the workload, we also presented a simple financial analysis of the impacts of such an organisational measures.

9.5.2 The DSR outputs from this work

According to March and Smith (1995), a DSR creates artefacts, methods, artefacts or instantiations to serve human purposes (PCA's requirement to be able to manage PHC centres in our case). In this sense, we have three DSR output elements from our work.

1. The theoretical constructs: According to March and Smith (1995), in DSR the concepts and the relationship between them is referred to a theoretical concepts. Care pathways are formalised standard guidelines to address care needs due to a specific health condition. In this thesis, we consider LTCs. As presented in Chapter 2, applying care pathways to individual patients may be referred to as care plans or plans-of-care. So, in this thesis, we make a clear distinction in using these terms. The box below highlights our definitions of these terms in the context of this thesis. We also define a clear relationship (instantiation, abstraction and aggregation) between them (see Section 3.1).

Care pathway: focuses on a specific LTC management needs.
Care plan: focuses on a specific LTC's management need in an individual.
Plan-of-care: addresses multi-morbidity management needs in a patient.

2. A process or a method: During a patient visit, our rules determine subsequent recalls for the patient by considering the disease-specific care plans (then) applicable to the patient, then merging them into an individual plan of care for the patient. In a later stage, we aggregate them as a population level workload. We call this process of unfolding care visit by visit for a patient as the encounter-based unfolding plan-of-care process.

3. The artefact: With our encounter-based unfolding plan-of-care process, during a patient visit, our ABP-WPM decides subsequent recalls for the patient. We were supplied with an anonymised dataset of a medical practice in New Zealand as part of the collaboration on this work. In our workload initialisation and prediction process, we compensated the shortcomings of the dataset by defining a few assumptions and definitions. We acknowledge that some of these assumptions maybe practice specific. We seeded our workload prediction model with care pathways as rules, and the preprocessed version of the anonymised patient dataset shared with us. The care pathways then undergoes the encounter-based plan-of-care process for each patient in the patient dataset allowing our workload model to simulate patient recalls. When the simulation ends, these recalls are aggregated on a weekly basis as the workload for the simulation period.

9.5.3 Further contributions

Apart from the above mentioned what-if scenarios, we also considered a scenario where an external factor (influenza outbreak) is expected to impact adversely on the workload. We showed that having an understanding of the upcoming workload, the practice can then take measures (as described in Section 6.3) to make slots available (from LTC appointments) to meet the workload due to an influenza outbreak. We also discussed financial analysis in these various what-if scenarios.

We also investigated using Bayesian inference to learn patient's prescription renewal dates from their full history of visits and incorporate in our simulation the variation in patient visits (with respect to their prescription renewal date) based on each patient's visit pattern. We showed that using Bayesian inference better aligned the predicted workload with the actual visit pattern of patients. However, there are a few challenges

that hindered using Bayesian inference to analyse our what-if scenarios. These are left for future work.

We claim that this encounter-based unfolding plan-of-care contributes to usage of care pathways in health domain. We discussed how the care pathways with patient data unfolds to disease specific care plans which then are used to create individualised plans-of-care during a patient visit. We also presented the iterative cycles of the design science research methodology (DSRM) followed to develop our adaptable best practice based workload prediction model (ABP-WPM).

9.6 Limitations and future work

As a proof-of-concept model, we have implemented care plans for three LTCs (namely CKD, diabetes and hypertension), and if this model is to be adopted at an organisation, all the care pathways need to be codified.

As presented in Section 3.2.2, in our final phase of development, we collected feedback from executive members of WellSouth Primary Health Network in Otago[1] including chief executive officer (CEO), chief information officer (CIO), nursing director, and the medical director, the general manager of primary care, Southern District Health Board, as well as the primary care facilitator, Southern Alliance[2]. The CEO highlighted that having an understanding of the impacts of various scenarios can help in the workforce development of a practice (which is currently lacking). One point that was highlighted was a comparison of their initiative to manage patients who are at high risk for hospitalisation where they try to stratify patients based on their LTC needs similar to recommended by the KP-risk stratification model (Pines et al., 2015). They too agree that there is no mechanism to identify acute and LTC workload and so, they do not consider acute care under this program.

In our work, we focused on the appointments to address the clinical complexities of LTCs in a patient. The nursing director of WellSouth Primary Health Network highlighted that social complexities also contribute to workload from LTC patients. Codifying social complexities of a patient or a group of patients is more complicated due to lack of a consensus definition of "social"-complexity of a patient (Mount et al., 2015).

In our work, we do a data pre-processing step in order to improve the underlying

[1]This is the Primary Health Organisation for Otago and South Island of New Zealand.
[2]Southern Alliance is a health insurance provider.

data model to have the capability to support population-level workload analysis. In that process of information inference, we identified what is required in a data model to support a population-level workload analysis. We developed the patient information model presented in Chapter 8 using the Web Ontology Language (OWL) tool. As a future work, we would extend the patient information model to provide the capability to reason over data. This would give our patient information model the capability to answer questions such as how many diabetic patients are currently seen by a specific GP, and out of them how many are at a high risk of CKD; or, how many young patients (e.g., $age < 30$) need to be supported with a group fitness class to manage their diabetes.

We demonstrated that we can learn using Bayesian inference the variation in patient visit with respect to their medication renewal period. Currently, we use all the historic visits of a patient. In our simulation, with the learnt probability of variation, we decide the next visit for a patient, which could be a weekend. Our rules then push the patient visits to a working day. As a future work, we could consider avoiding the chances of a patient visiting on a weekend, so that when a sample gives the offset from a reference date, it would refer to number of working days since the reference date.

In our work when defining what-if scenarios, the scenarios were encoded in the .drl rule formats. This means if a practitioner needs the autonomy to define new what-if scenarios, the practitioner requires some basic knowledge of Drools and its rule formats. In future, this work can be extended to develop a graphical user interface (GUI) for specifying various user specific what-if scenarios.

This work currently considers a time frame of a year for its simulation period. This work could be extended to consider the variability in plan-of care for a longer simulation period (e.g., 5 years). This would also require to consider patient outcomes e.g., National Kidney Foundation (2015) highlights the rate of kidney deterioration over 5 year period based on patient features such as ethnicity and age.

9.7 Longer term impact of this line of work

The findings from this work can be utilised to plan organisational-level policies so that most of their chronic patients would benefit. This work can also help the medical practice to strengthen their workforce based on their changing cohort of chronic patients. This work with further add-ins such as sending alerts for medication due to patient's smart phones, and emailing patients reminding them of their upcoming appointment,

can help medical practices systematise care for their patients. In a wider scope, knowing the upcoming workload can also help the funding and primary health organisations formulate schemes to address practice level needs, in addition to the 'one size fits all' strategy at a national level.

Conclusion

This thesis forms a initial work to understand the workload from the LTC population at a medical practice. This work has take different paths in future. An immediate work is to consider all respective care pathways for chronic conditions than assuming a general pathway for patients other than diabetes, CKD and hypertension. Another path is to consider the variability in plan-of care for a longer simulation period (e.g., 5 years). The findings from this work can be utilised to plan organisational-level policies that most of their chronic patients would benefit. This work can also help the medical practice to strengthen their workforce based on their changing cohort of chronic patients. Similar to CCM is adopted in many other care programs like CM+ (care management plus for chronic patients), this thesis with a further add-ins can help medical practices systematise care for their patients.

References

Abdel-Aal, R. and Mangoud, A. (1998). Modeling and forecasting monthly patient volume at a primary health care clinic using univariate time-series analysis. *Computer Methods and Programs in Biomedicine*, 56(3):235–247.

Agency for Healthcare Research and Quality (2014a). Chapter 2. What is Care Coordination? URL: `www.ahrq.gov/professionals/prevention-chronic-care/improve/coordination/atlas2014/chapter2.html`.

Agency for Healthcare Research and Quality (2014b). Clinical Guidelines and Recommendations. URL: `www.ahrq.gov/professionals/clinicians-providers/guidelines-recommendations/index.html`.

Åhgren, B. (2003). Chain of care development in Sweden: results of a national study. *International Journal of Integrated Care*, 3. PMC1483939.

Alther, M. and Reddy, C. K. (2015). Chapter 19: Clinical Decision Support Systems. In Reddy, C. K. and Aggarwal, C. C., editors, *Healthcare Data Analytics*, DataMining and Knowledge Discovery Series, pages 625–653. A Chapman and Hall Book, CRC Press. ISBN: 978-1-4822-3211-0.

Amir, O., Grosz, B. J., Gajos, K. Z., Swenson, S. M., and Sanders, L. M. (2015). From care plans to care coordination: Opportunities for computer support of teamwork in complex healthcare. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1419–1428. ACM. DOI: 10.1145/2702123.2702320.

Anonymous (2006a). Jargon Buster - Working in Partnership Programme (WiPP). URL: `https://www.gponline.com/jargon-buster-working-partnership-programme-wipp/article/588276`.

Anonymous (2006b). Workload analysis tool piloted. *Practice Nurse: The Journal for Nurses in General Practice; Philadelphia*, 32(10):7. ISSN: 09536612.

Anooj, P. K. (2012). Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. *Journal of King Saud University - Computer and Information Sciences*, 24(1):27–40. DOI: 10.1016/j.jksuci.2011.09.002.

Archenaa, J. and Anita, E. A. M. (2015). A Survey of Big Data Analytics in Healthcare and Government. *Procedia Computer Science*, 50:408–413. DOI: 10.1016/j.procs.2015.04.021.

Ash, A. S., Ellis, R. P., Pope, G. C., Ayanian, J. Z., Bates, D. W., Burstin, H., Iezzoni, L. I., MacKay, E., and Yu, W. (2000). Using Diagnoses to Describe Populations and Predict Costs. *Health Care Financing Review*, 21(3):7–28. PMID: 11481769.

Aysola, J., Rhodes, K. V., and Polsky, D. (2015). Patient-centered Medical Homes and Access to Services for New Primary Care Patients. *Medical Care*, 53(10):857–862. DOI: 10.1097/MLR.0000000000000412.

Azadmanjir, Z., Safdari, R., Ghazisaeedi, M., Mokhtaran, M., and Mohammad, E. K. (2017). A Three-Phase Decision Model of Computer-Aided Coding for the Iranian Classification of Health Interventions (IRCHI). *Acta Informatica Medica*, 25(2):88–93. DOI: 10.5455/aim.2017.25.88-93.

Babiker, A., El Husseini, M., Al Nemri, A., Al Frayh, A., Al Juryyan, N., Faki, M. O., Assiri, A., Al Saadi, M., Shaikh, F., and Al Zamil, F. (2014). Health care professional development: Working as a team to improve patient care. *Sudanese Journal of Paediatrics*, 14(2):9–16. PMCID: PMC4949805.

Baskerville, R., Pries-Heje, J., and Venable, J. (2009). Soft design science methodology. In *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, page 9. ACM. DOI:10.1145/1555619.1555631.

Baud, R. (2003). *The New Navigators: From Professionals to Patients : Proceedings of MIE2003*. IOS Press. Google-Books-ID: i0KEzajbRUIC.

Bauman, A. E., Fardy, H. J., and Harris, P. G. (2003). Getting it right: why bother with patient-centred care? *The Medical Journal of Australia*, 179(5):253–256. PMID: 12924973.

Best Practice Advocacy Centre New Zealand (1997). bestpractice Decision Support - About BPAC Inc. URL: http://www.bestpractice.net.nz/home_about.php.

Best Practice Advocacy Centre New Zealand (2012). Care pathways for long-term conditions. *Best Practice Journal*, 47:38–41. URL: http://www.bpac.org.nz/BPJ/2012/october/carepathways.aspx.

Bilello, L. A., Hall, A., Harman, J., Scuderi, C., Shah, N., Mills, J. C., and Samuels, S. (2018). Key attributes of patient centered medical homes associated with patient activation of diabetes patients. *BMC family practice*, 19(1):4. DOI: 10.1186/s12875-017-0704-3.

Bodenheimer, T., Chen, E., and Bennett, H. D. (2009). Confronting The Growing Burden Of Chronic Disease: Can The U.S. Health Care Workforce Do The Job? *Health Affairs*, 28(1):64–74. DOI: 10.1377/hlthaff.28.1.64.

Bodenheimer, T., MacGregor, K., and Sharifi, C. (2005). Helping patients manage their chronic conditions. California Health Care Foundation.

Bodenheimer, T., Wagner, E. H., and Grumbach, K. (2002a). Improving Primary Care for Patients With Chronic Illness. *JAMA*, 288(14):1775–1779. DOI:10.1001/jama.288.14.1775.

Bodenheimer, T., Wagner, E. H., and Grumbach, K. (2002b). Improving Primary Care for Patients With Chronic Illness: The Chronic Care Model, Part 2. *JAMA*, 288(15):1909–1914. DOI:10.1001/jama.288.15.1909.

Borgmeyer, C. (2013). 2014 Hypertension Guideline Stands to Simplify Treatment, Says Expert. URL: https://www.aafp.org/news/health-of-the-public/20131218hypertensiongdln.html.

Brailsford, S. and Hilton, N. (2001). A Comparison of Discrete Event Simulation and System Dynamics for Modelling Healthcare Systems. pages 1–17. URL: https://eprints.soton.ac.uk/35689/1/glasgow_paper.pdf.

Breton, M., Lvesque, J.-F., Pineault, R., and Hogg, W. (2011). Primary Care Reform: Can Quebec's Family Medicine Group Model Benefit from the Experience of Ontario's Family Health Teams? *Healthcare Policy*, 7(2):e122–e135.

Brown, S., Lhussier, M., Dalkin, S. M., and Eaton, S. (2018). Care Planning: What Works, for Whom, and in What Circumstances? A Rapid Realist Review. *Qualitative Health Research*. DOI: 10.1177/1049732318768807.

Burt, J., Rick, J., Blakeman, T., Protheroe, J., Roland, M., and Bower, P. (2014). Care plans and care planning in long-term conditions: a conceptual model. *Primary health care research & development*, 15(04):342–354. DOI: 10.1017/S1463423613000327.

Busetto, L., Luijkx, K., Calciolari, S., Gonzlez Ortiz, L. G., and Vrijhoef, H. J. M. (2017). Exploration of workforce changes in integrated chronic care: Findings from an interactive and emergent research design. *PloS One*, 12(12):e0187468. DOI: 10.1371/journal.pone.0187468.

Calcaterra, D., Di Modica, G., Tomarchio, O., and Romeo, P. (2018). A clinical decision support system to increase appropriateness of diagnostic imaging prescriptions. *Journal of Network and Computer Applications*, 117:17–29. WOS:000438005400003.

Calvan, O., Penny, D., Jean, D., Fountaine, T., and Raihi, F. (2011). Using care pathways to improve health systems. *Health International*, 11.

Campbell-Scherer, D. (2010). Multimorbidity: a challenge for evidence-based medicine. *Evidence-Based Medicine*, 15(6):165–166. DOI: 10.1136/ebm1154.

Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., Morton, S. C., and Shekelle, P. G. (2006). Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Annals of Internal Medicine*, 144(10):742–752.

Chute, C. G. (2000). Clinical Classification and Terminology: Some History and Current Observations. *Journal of the American Medical Informatics Association*, 7(3):298–303. DOI: 10.1136/jamia.2000.0070298.

Clinica (2018). Clinica Family Health Home. URL: `https://www.clinica.org/`.

Connell, J., Carlton, J., Grundy, A., Buck, E. T., Keetharuth, A. D., Ricketts, T., Barkham, M., Robotham, D., Rose, D., and Brazier, J. (2018). The importance of content and face validity in instrument development: lessons learnt from service users when developing the Recovering Quality of Life measure (ReQoL). *Quality of Life Research*, 27(7):1893–1902. WOS:000435128700021.

Dahabreh, I. J., Chan, J. A., Earley, A., Moorthy, D., Avendano, E. E., Trikalinos, T. A., Balk, E. M., and Wong, J. B. (2017). Chapter 4, A Review of Validation and Calibration Methods for Health Care Modeling and Simulation. In *Modeling and Simulation in the Context of Health Technology Assessment: Review of Existing Guidance, Future Research Needs, and Validity Assessment*. Agency for Healthcare Research and Quality (US).

Dale, M. C. (2015). Turning silver to gold: Policies for an ageing population. URL:`https://cdn.auckland.ac.nz/assets/business/about/our-research/research-institutes-and-centres/RPRC/WorkingPaper/WP\%202014-2\%20LTC\%20costs\%20FINAL.pdf`.

Davidsson, P. (2001). Multi Agent Based Simulation: Beyond Social Simulation. In Moss, S. and Davidsson, P., editors, *Multi-Agent-Based Simulation*, Lecture Notes in Computer Science, pages 97–107. Springer Berlin Heidelberg.

Davis, R., Buchanan, B., and Shortliffe, E. (1977). Production rules as a representation for a knowledge-based consultation program. *Artificial Intelligence*, 8(1):15–45. DOI: 10.1016/0004-3702(77)90003-0.

Davy, C., Bleasel, J., Liu, H., Tchan, M., Ponniah, S., and Brown, A. (2015). Effectiveness of chronic care models: opportunities for improving healthcare practice and health outcomes: a systematic review. *BMC health services research*, 15:194. DOI: 10.1186/s12913-015-0854-8.

De Bleser, L., Depreitere, R., De Waele, K., Vanhaecht, K., Vlayen, J., and Sermeus, W. (2006). Defining pathways. *Journal of Nursing Management*, 14(7):553–563. DOI: 10.1111/j.1365-2934.2006.00702.x.

Decouttere, C. and Vandaele, N. (2014). A Broader View on Health Care System Design and Modelling. In *Proceedings of the International Conference on Health Care Systems Engineering*, pages 215–225. Springer, Cham. DOI: 10.1007/978-3-319-01848-5_17.

Dee Mangin (2012). Adherence to evidence-based guidelines is the key to improved health outcomes for general practice patients - the no case. *Journal of Primary Health Care*, 4(2):158–160.

Demski, H., Garde, S., and Hildebrand, C. (2016). Open data models for smart health interconnected applications: the example of openEHR. *BMC Medical Informatics and Decision Making*, 16(137). DOI: 10.1186/s12911-016-0376-2.

Dennis, S. M., Zwar, N., Griffiths, R., Roland, M., Hasan, I., Davies, G. P., and Harris, M. (2008). Chronic disease management in primary care: from evidence to policy. *Medical Journal of Australia*, 188:S53–S56.

Department of Health & Human Services (2016). Victoria's International Health Strategy 2016-2020. URL: `https://www2.health.vic.gov.au:443/about/health-strategies/international-health-strategy`.

Department of Health Nunavut (2018). Health Strategies of Government of Nunavut. URL: `https://www.gov.nu.ca/health/information/health-strategies-0`.

Devananda, M., Cranefield, S., Winikoff, M., and Lloyd, H. (2017). Workload prediction model of a primary health centre. In *25th European Conference on Information Systems (ECIS)*, pages 1192–1204, Guimares, Portugal.

DHS Primary Health Branch Victoria (2008). Revised Chronic Disease Management Program Guidelines for Primary Care Partner-ships and Primary Health Care Services. Technical report, DHS Primary Health Branch Victoria.

Downing, A., Rudge, G., Cheng, Y., Tu, Y.-K., Keen, J., and Gilthorpe, M. S. (2007). Do the UK government's new Quality and Outcomes Framework (QOF) scores adequately measure primary care performance? A cross-sectional survey of routine healthcare data. *Bmc Health Services Research*, 7:166. WOS:000251532800001.

du Vaure, C. B., Ravaud, P., Baron, G., Barnes, C., Gilberg, S., and Boutron, I. (2016). Potential workload in applying clinical practice guidelines for patients with chronic conditions and multimorbidity: a systematic analysis. *Bmj Open*, 6(3):e010119. WOS:000374052300067.

Eatock, J., Lord, J., Trapero-Bertran, M., and Anagnostou, A. (2015). Discrete Event Simulation of Whole Care Pathways to Estimate Cost-Effectiveness in Clinical Guidelines. In *2015 Winter Simulation Conference (wsc)*, pages 1447–1458. Ieee, New York. WOS:000399133901033.

Fernandez-Millan, R., Medina-Merodio, J.-A., Barchino Plata, R., Martinez-Herraiz, J.-J., and Gutierrez-Martinez, J.-M. (2015). A Laboratory Test Expert System

for Clinical Diagnosis Support in Primary Health Care. *Applied Sciences-Basel*, 5(3):222–240. WOS:000362644500005.

Fetherstonhaugh, D., Nay, R., and Winbolt, M. (2013). Evidence-based Health Care. In *Research Methods in Health: Foundations for Evidence Based Practice*, pages 251–65. Oxford University Press, Melbourne. ISBN: 978-0-19-552862-6.

Fischer, F., Lange, K., Klose, K., Greiner, W., and Kraemer, A. (2016). Barriers and Strategies in Guideline ImplementationA Scoping Review. *Healthcare*, 4(3):36. DOI: 10.3390/healthcare4030036.

Forbes, L. J., Marchand, C., Doran, T., and Peckham, S. (2017). The role of the Quality and Outcomes Framework in the care of long-term conditions: a systematic review. *The British Journal of General Practice*, 67(664):e775–e784.

Fox, J., Alabassi, A., Patkar, V., Rose, T., and Black, E. (2006). An ontological approach to modelling tasks and goals. *Computers in biology and medicine*, 36(7):837–856.

Fox, J., Patkar, V., Chronakis, I., and Begent, R. (2009). From practice guidelines to clinical decision support: closing the loop. *Journal of the Royal Society of Medicine*, 102(11):464–473.

Gacenga, F., Cater-Steel, A., Toleman, M., and Tan, W.-G. (2012). A Proposal and Evaluation of a Design Method in Design Science Research. *The Electronic Journal of Business Research Methods*, 10(2):89–100.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3):515–534. DOI: 10.1214/06-BA117A.

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4):1360–1383. arXiv: 0901.4011.

Glasgow, R. E., Orleans, C. T., Wagner, E. H., Curry, S. J., and Solberg, L. I. (2001). Does the Chronic Care Model Serve Also as a Template for Improving Prevention? *The Milbank Quarterly*, 79(4):579–612. DOI:10.1111/1468-0009.00222.

Goertzel, G. (1969). Clinical Decision Support System. *Annals of the New York Academy of Sciences*, 161(2):689–693. DOI:10.1111/j.1749-6632.1969.tb34100.x.

Government of Canada Health (2004). Strategies and Initiatives. URL: https://www.canada.ca/en/health-canada/corporate/about-health-canada/activities-responsibilities/strategies-initiatives.html.

Grant, R. and Greene, D. (2012). The Health Care Home Model: Primary Health Care Meeting Public Health Goals. *American Journal of Public Health*, 102(6):1096–1103. DOI: 10.2105/AJPH.2011.300397.

Green, L. (2013). Queueing Analysis in Health Care. In *Patient Flow*, International Series in Operations Research & Management Science, pages 361–384. Springer, Boston, MA. ISBN: 978-1-4614-9511-6 978-1-4614-9512-3.

Green, L. V. and Savin, S. (2008). Reducing Delays for Medical Appointments: A Queueing Approach. *Operations Research*, 56(6):1526–1538. DOI: 10.1287/opre.1080.0575.

Green Cross Health (2017). The Doctors: Explaining primary health in NZ. URL: https://thedoctors.co.nz/Explaining-primary-health-in-NZ.

Grundy, P. H. and Hodach, R. J. (2016). Health ITs Essential Role in the Patient-Centered Medical Home and Practice-Based Population Health Management. In Weaver, C. A., Ball, M. J., Kim, G. R., and Kiel, J. M., editors, *Healthcare Information Management Systems*, Health Informatics, pages 243–255. Springer International Publishing. DOI: 10.1007/978-3-319-20765-0_15.

Günal, M. M. and Pidd, M. (2010). Discrete event simulation for performance modelling in health care: a review of the literature. *Journal of Simulation*, 4(1):42–51.

Hall, R. (2012). Chapter 1: Matching Healthcare Resources to Patient Needs. In Hall, R., editor, *Handbook of Healthcare System Scheduling*, number 168 in International Series in Operations Research & Management Science. Springer Science+Business Media, LLC 2012. ISBN: 978-1-4614-1734-7_1.

Ham, C. (2010). The ten characteristics of the high-performing chronic care system. *Health Economics, Policy, and Law*, 5(Pt 1):71–90. DOI: 10.1017/S1744133109990120.

Harris, M., Lawn, S. J., Morello, A., Battersby, M. W., Ratcliffe, J., McEvoy, R. D., and Tieman, J. J. (2017). Practice change in chronic conditions care: an appraisal of

theories. *BMC Health Services Research*, 17(1):170. DOI: 10.1186/s12913-017-2102-x.

Harris, M. F. and Zwar, N. A. (2007). Care of patients with chronic disease: the challenge for general practice. *Medical Journal of Australia*, 187(2):104–107.

Hawkins, M. R. and Novak, T. (2011). Guided Care Program at Johns Hopkins HealthCare Wins CMSAs 2011 Case Management Research Award. *CMSA Today*, (4). URL: http://www.nxtbook.com/nxtbooks/naylor/CMSQ0211/index.php?startid=14\#/14.

Haynes, B. and Haines, A. (1998). Barriers and bridges to evidence based clinical practice. *BMJ : British Medical Journal*, 317(7153):273–276.

He, W., Goodkind, D., and Kowal, P. (2016). An aging world:2015. Technical report, U.S. Government Publishing Office, Washington (DC). URL: https://www.census.gov/content/dam/Census/library/publications/2016/demo/p95-16-1.pdf.

HealthPartners (2018). Learn how HealthPartners is transforming the delivery of health care and insurance. URL: https://www.healthpartners.com/hp/about/index.html.

Hefford, B. (2006). Future delivery of community based health and care services. Technical report, Capital & Coast District Health Board, New Zealand. URL: http://www.achc.org.co/hospital360/tendencias_emergentes/Atencion\%20_en_salud_basada_en_la_comunidad.pdf.

Heroman, W. M., Davis, C. B., and Farmer, K. L. (2012). Demand Forecasting and Capacity Management in Primary Care. *Physician Executive; Tampa*, 38(1):30–4.

Hevner, A. R. (2007). A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, 19(2). URL: http://aisel.aisnet.org/sjis/vol19/iss2/4.

Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design Science in Informations Research. *MIS Quarterly*, 28(1):75–105.

HL7 (2016). HL7 Standards Product Brief - HL7 Version 3 Product Suite. URL:https://www.hl7.org/implement/standards/product_brief.cfm?product_id=186.

Iacobucci, G. (2016). GP appointments should be 15 minutes long, says BMA. *British Medical Journal*, 354:i4709. DOI: 10.1136/bmj.i4709.

Ibrahim, M. A., Savitz, L. A., Carey, T. S., and Wagner, E. H. (2001). Population-based health principles in medical and public health practice. *Journal of public health management and practice: JPHMP*, 7(3):75–81. PMID: 11338089.

Indian Health Service (2014). Care Management Event Tracking (CMET). URL: `https://www.ihs.gov/icare/cmet/`.

Institute for Quality and Efficiency in Health Care (2016). What are disease management programs (DMPs)? *Informed Health*. URL:`https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0072596/`.

Institute of Medicine (US) Committee (2002). *Understanding Population Health and Its Determinants*. National Academies Press (US), assuring the health of the public in the 21st century edition.

Institute of Medicine (US) Committee on Quality of Health Care in America (2001). *Crossing the Quality Chasm: A New Health System for the 21st Century*. National Academies Press (US), Washington (DC). ISBN: 978-0-309-07280-9.

Institute of Medicine (US) Committee on the Future of Primary Care (1996). Defining Primary Care. In Donaldson, M. S., Yordy, K. D., Lohr, K. N., and Vanselow, N. A., editors, *Primary Care: America's Health in a New Era*. National Academies Press (US). ISBN: 0-309-05399-4.

Jackson, G. L., Powers, B. J., Chatterjee, R., Bettger, J. P., Kemper, A. R., Hasselblad, V., Dolor, R. J., Irvine, R. J., Heidenfelder, B. L., Kendrick, A. S., Gray, R., and Williams, J. W. (2013). Improving patient care. The patient centered medical home. A Systematic Review. *Annals of Internal Medicine*, 158(3):169–178. DOI: 10.7326/0003-4819-158-3-201302050-00579.

Johnson, S. (1997). *Pathways of Care*. Wiley. ISBN: 978-0-632-04076-6.

Jordan, K., Ong, B. N., and Croft, P. (2003). Previous consultation and self reported health status as predictors of future demand for primary care. *Journal of Epidemiology & Community Health*, 57(2):109–113. DOI: 10.1136/jech.57.2.109.

Jorgensen, T. S., Skougaard, M., Taylor, P. C., Asmussen, H. C., Lee, A., Klokker, L., Svejstrup, L., Mountian, I., Gudbergsen, H., and Kristensen, L. E. (2018). The Parker Model: Applying a Qualitative Three-Step Approach to Optimally Utilize Input from Stakeholders When Introducing New Device Technologies in the Management of Chronic Rheumatic Diseases. *Patient-Patient Centered Outcomes Research*, 11(5):515–526. WOS:000444011000006.

Jun, J. B., Jacobson, S. H., and Swisher, J. R. (1999). Application of discrete-event simulation in health care clinics: A survey. *Journal of the Operational Research Society*, 50(2):109–123.

Kahn, M. G., Fagan, L. M., and Tu, S. (1991). Extensions to the Time-Oriented Database Model to Support Temporal Reasoning in Medical Expert Systems. *Methods of Information in Medicine*, 30(1):4–14. DOI: 10.1055/s-0038-1634816.

Kane, R. L., Priester, R., and Totten, A. M. (2005). *Meeting the challenge of chronic illness*. Baltimore : Johns Hopkins University Press. ISBN: 978-0-8018-8209-8.

Kim, J.-S., Gao, X., and Rzhetsky, A. (2018). RIDDLE: Race and ethnicity Imputation from Disease history with Deep LEarning. *Plos Computational Biology*, 14(4):e1006106. WOS:000432169600043.

Kindig, D. and Stoddart, G. (2003). Models for Population Health: What Is Population Health? *American Journal of Public Health*, 93(3):380–383. DOI: 10.2105/AJPH.93.3.380.

Kohli, R. and Tan, S. S.-L. (2016). Electronic Health Records: How Can IS Researchers Contribute to Transforming Healthcare? *Mis Quarterly*, 40(3):553–573.

Konrad, R., Tang, C., Shiner, B., and Watts, B. V. (2017). Workforce design in primary care-mental health integration: a case study at one veterans affairs medical center. *Health Systems*, 6(2):148–160. WOS:000411321900006.

Krittanawong, C., Bomback, A. S., Baber, U., Bangalore, S., Messerli, F. H., and Tang, W. H. W. (2018). Future Direction for Using Artificial Intelligence to Predict and Manage Hypertension. *Current Hypertension Reports*, 20(9):75. WOS:000437804900003.

Kuo, K.-L. and Fuh, C.-S. (2009). A Rule-Based Clinical Decision Model to Support Interpretation of Multiple Data in Health Examinations. *Journal of Medical Systems*. DOI: 10.1007/s10916-009-9413-3.

Law, A. M. and Kelton, W. D. (1991). *Simulation Modeling and Analysis*. McGraw-Hill, New York.

Leukel, J., Mueller, M., and Sugumaran, V. (2014). The State of Design Science Research within the BISE Community: An Empirical Investigation. In *Thirty Fifth International Conference on Information Systems*, page 15, Auckland.

Levey, A. S., Stevens, L. A., Schmid, C. H., Zhang, Y. L., Castro III, A. F., Feldman, H. I., Kusek, J. W., Eggers, P., Van Lente, F., Greene, T., and Coresh, J. (2009). A New Equation to Estimate Glomerular Filtration Rate. *Annals of Internal Medicine*, 150(9):604–612.

Lhotska, L., Marik, V., and Vlcek, T. (2001). Medical applications of enhanced rule-based expert systems. *International Journal of Medical Informatics*, 63(1):61–75. DOI: 10.1016/S1386-5056(01)00172-1.

Lin, Y., Huang, S., Simon, G. E., and Liu, S. (2018). Data-based Decision Rules to Personalize Depression Follow-up. *Scientific Reports*, 8:5064. WOS:000428034400015.

Luke, D. A., Sarli, C. C., Suiter, A. M., Carothers, B. J., Combs, T. B., Allen, J. L., Beers, C. E., and Evanoff, B. A. (2018). The Translational Science Benefits Model: A New Framework for Assessing the Health and Societal Benefits of Clinical and Translational Sciences: Translational Science Benefits Model. *Clinical and Translational Science*, 11(1):77–84.

Mabotuwana, T. and Warren, J. (2010). ChronoMedItA computational quality audit framework for better management of patients with chronic conditions. *Journal of Biomedical Informatics*, 43(1):144–158.

Maher, D., Harries, A. D., Zachariah, R., and Enarson, D. (2009). A global framework for action to improve the primary care response to chronic non-communicable diseases: a solution to a neglected problem. *BMC Public Health*, 9(1):355. DOI: 10.1186/1471-2458-9-355.

March, S. T. and Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, 15(4):251–266. DOI:10.1016/0167-9236(94)00041-2.

Marshall, J., Chahin, A., and Rush, B. (2016). Review of Clinical Databases. In *Secondary Analysis of Electronic Health Records*, pages 9–16. Springer International Publishing. DOI: 10.1007/978-3-319-43742-2_2.

Massimi, A., Vito, C. D., Brufola, I., Corsaro, A., Marzuillo, C., Migliara, G., Rega, M. L., Ricciardi, W., Villari, P., and Damiani, G. (2017). Are community-based nurse-led self-management support interventions effective in chronic patients? Results of a systematic review and meta-analysis. *PLOS ONE*, 12(3):e0173617. DOI: 10.1371/journal.pone.0173617.

Mathers, N., Roberts, S., Hodkinson, I., and Karet, B. (2011). Care planning: improving the lives of people with long term conditions. *London: Royal College of General Practitioners*.

Mays, N. (2013). Reorienting the New Zealand health care system to meet the challenge of long-term conditions in a fiscally constrained environment. URL:`http://www.victoria.ac.nz/sacl/centres-and-institutes/cpf/publications/pdfs/Nick-Mays-Revised-Conference-Paper-Jan-2013-website-version.pdf`.

McPhail, S. M. (2016). Multimorbidity in chronic disease: impact on health care resources and costs. *Risk Management and Healthcare Policy*, 9:143–156. DOI: 10.2147/RMHP.S97248.

MedTech Limited (2013). Medtech32 - New Zealand. URL: `http://www.medtechglobal.com/nz/`.

Mickan, S. M. and Rodger, S. A. (2005). Effective Health Care Teams: A model of six characteristics developed from shared perceptions. *Journal of Interprofessional Care*, 19(4):358–370. DOI: 10.1080/13561820500165142.

Ministry of Health NZ (2000). The future shape of primary health care. URL: `http://www.moh.govt.nz/notebook/nbbooks.nsf/ea5ef2c0e4ab8ac485256caa0065e3eb/15de5a57c07f857e4c2568b100731ab7/\$FILE/primarycare.pdf`.

Ministry of Health NZ (2014). Primary health care providers. URL:http://www.health.govt.nz/our-work/primary-health-care/about-primary-health-organisations/primary-health-care-providers.

Ministry of Health NZ (2016a). Living Well with Diabetes: 2015-2020. URL:http://www.health.govt.nz/our-work/diseases-and-conditions/diabetes/living-well-diabetes-2015-2020.

Ministry of Health NZ (2016b). More Heart and Diabetes Checks Evaluation. URL:http://www.health.govt.nz/publication/more-heart-and-diabetes-checks-evaluation.

Ministry of Health NZ (2016c). New Zealand Burden of Diseases, Injuries and Risk Factors Study. URL:http://www.health.govt.nz/nz-health-statistics/health-statistics-and-data-sets/new-zealand-burden-diseases-injuries-and-risk-factors-study.

Ministry of Health NZ (2016d). New Zealand Health Strategy:Future direction. URL: http://www.health.govt.nz/system/files/documents/publications/new-zealand-health-strategy-futuredirection-2016-apr16.pdf.

Ministry of Health NZ (2016e). Primary Health Care Services Funding and Contracting. URL: https://www.health.govt.nz/system/files/documents/pages/primary-health-care-funding-2016.pdf.

Ministry of Health NZ (2017a). About primary health organisations. URL: https://www.health.govt.nz/our-work/primary-health-care/about-primary-health-organisations.

Ministry of Health NZ (2017b). Care Plus. URL:http://www.health.govt.nz/our-work/primary-health-care/primary-health-care-subsidies-and-services/care-plus.

Ministry of Health NZ (2017c). Long-term conditions. URL:http://www.health.govt.nz/our-work/diseases-and-conditions/long-term-conditions.

Ministry of Health NZ (2017d). Mori health models. URL: https://www.health.govt.nz/our-work/populations/maori-health/maori-health-models.

Ministry of Health NZ (2017e). Mori health models Te Pae Mahutonga. URL: https://www.health.govt.nz/our-work/populations/maori-health/maori-health-models/maori-health-models-te-pae-mahutonga.

Ministry of Health NZ (2017f). Mori health models Te Whare Tapa Wh. URL: https://www.health.govt.nz/our-work/populations/maori-health/maori-health-models/maori-health-models-te-whare-tapa-wha.

Ministry of Health NZ (2017g). Mori health models Te Wheke. URL: https://www.health.govt.nz/our-work/populations/maori-health/maori-health-models/maori-health-models-te-wheke.

Ministry of Health NZ (2017h). *New Zealand Health Research Strategy 20172027: Summary of Submissions and Consultation*. URL: http://www.health.govt.nz/publication/new-zealand-health-research-strategy-2017-2027-summary-submissions-/and-consultation.

Ministry of Health NZ (2017i). Primary health care. URL: https://www.health.govt.nz/our-work/primary-health-care.

Ministry of Health NZ (2017j). SNOMED CT Implementation in New Zealand. URL:http://www.health.govt.nz/nz-health-statistics/classification-and-terminology/new-zealand-snomed-ct-national-release-centre/snomed-ct-implementation-new-zealand.

Ministry of Health NZ (2017k). Workforce issues: public health, health jobs, health issues - Public Health Workforce Development. URL: http://www.publichealthworkforce.org.nz/health-jobs-workforce-health-issues_167.aspx.

Minutolo, A., Esposito, M., and De Pietro, G. (2017). Optimization of rule-based systems in mHealth applications. *Engineering Applications of Artificial Intelligence*, 59:103–121. DOI: 10.1016/j.engappai.2016.12.007.

Montague, E. (2014). The promises and challenges of health information technology in primary health care. *Primary health care research & development*, 15(03):227–230. 10.1017/S1463423614000231.

Mount, J. K., Massanari, R. M., and Teachman, J. (2015). Patient Care Complexity as Perceived by Primary Care Physicians. *Famililes, Systems, & Health*, 33(2):137–145. DOI: 10.1037/fsh0000122.

Murphy, K. P. (2012). *Machine Learning: a Probabilistic Perspective*. MIT Press.

Murray, M. and Berwick, D. M. (2003). Advanced Access: Reducing Waiting and Delays in Primary Care. *JAMA*, 289(8):1035–1040. DOI:10.1001/jama.289.8.1035.

Mustapha, K., Gilli, Q., Frayret, J.-M., and Lahrichi, N. (2018). Agent-Based Modelling and Simulation Framework for Health Care. In Obaidat, M. S., Oren, T., and Merkuryev, Y., editors, *Simulation and Modeling Methodologies, Technologies and Applications, Simultech 2016*, volume 676, pages 171–197. Springer International Publishing Ag, Cham. WOS:000437042300011.

National Health Board NZ (2010). Trends in Service Design and New Models of Care - A Review. Technical report, Ministry of Health NZ, Wellington, New Zealand. URL: `http://www.moh.govt.nz/NoteBook/nbbooks.nsf/0/43C4C9001223BCBDCC257B10007AC681/\$file/trends-service-design-new-models-care-jul2010.pdf`.

National Kidney Foundation (2015). Glomerular Filtration Rate (GFR). URL:`https://www.kidney.org/atoz/content/gfr`.

NHS Digital (2017a). Read Codes - NHS Digital. URL: `https://digital.nhs.uk/article/1104/Read-Codes`.

NHS Digital (2017b). SNOMED CT implementation in primary care - NHS Digital. URL:`https://digital.nhs.uk/SNOMED-CT-implementation-in-primary-care`.

NICE (1999). NICE Pathways. URL:`https://www.nice.org.uk/about/what-we-do/our-programmes/about-nice-pathways`.

NZGG (2011). Guidance on the Management of Type 2 Diabetes. URL:`http://bth.diva-portal.org/smash/get/diva2:809928/FULLTEXT01.pdf`.

NZNO (2016). Primary Health Care Multi-Employer Collective Agreement. URL: `https://www.nzma.org.nz/\_\_data/assets/pdf\_file/0016/55402/PHC-MECA-2016-Final-Signed.pdf`.

O'Halloran, J., Miller, G. C., and Britt, H. (2004). Defining chronic conditions for primary care with ICPC-2. *Family Practice*, 21(4):381–386. DOI: 10.1093/fampra/cmh407.

O'Leary, P., Noll, J., and Richardson, I. (2013). A resource flow approach to modelling care pathways. In *International Symposium on Foundations of Health Informatics Engineering and Systems*, pages 41–58, Macau. Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-53956-5_4.

ONC (2016). Quality Data Model, Version 4.3. URL:`https://ecqi.healthit.gov/system/files/qdm\_4\_3\_508\_compliant.pdf`.

openEHR (2016). openEHR Information Model Release 1.0.3. URL: `http://www.openehr.org/releases/RM/Release-1.0.3/docs/ehr/ehr.html`.

Overington, J. D., Huang, Y. C., Abramson, M. J., Brown, J. L., Goddard, J. R., Bowman, R. V., Fong, K. M., and Yang, I. A. (2014). Implementing clinical guidelines for chronic obstructive pulmonary disease: barriers and solutions. *Journal of Thoracic Disease*, 6(11):1586–1596. DOI: 10.3978/j.issn.2072-1439.2014.11.25.

OWL Working Group (2012). OWL - Semantic Web Standards. URL: `https://www.w3.org/OWL/`.

Papageorgiou, E. I. (2011). A new methodology for Decisions in Medical Informatics using fuzzy cognitive maps based on fuzzy rule-extraction techniques. *Applied Soft Computing*, 11(1):500–513. DOI: 10.1016/j.asoc.2009.12.010.

Papatheodorou, K., Banach, M., Edmonds, M., Papanas, N., and Papazoglou, D. (2015). Complications of Diabetes. DOI: 10.1155/2015/189525.

Pawlak, Z. (1997). Rough set approach to knowledge-based decision support. *European Journal of Operational Research*, 99(1):48–57. DOI: 10.1016/S0377-2217(96)00382-7.

Payne, T. H., Galvin, M., Taplin, S. H., Austin, B., Savarino, J., and Wagner, E. H. (1995). Practicing population-based care in an HMO: evaluation after 18 months. *HMO practice*, 9(3):101–106. PMID: 10151092.

PayScale New Zealand (2018). People with Jobs as Physicians / Doctors Wages, Hourly Wage Rate. URL: `https://www.payscale.com/research/NZ/People\_with\_Jobs\_as\_Physicians\_\%2f\_Doctors/Hourly\_Rate`.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77.

Peleg, M. (2013). Computer-interpretable clinical guidelines: A methodological review. *Journal of Biomedical Informatics*, 46(4):744–763. DOI: 10.1016/j.jbi.2013.06.009.

Pestotnik, S. L., Classen, D. C., Evans, R. S., and Burke, J. P. (1996). Implementing Antibiotic Practice Guidelines through Computer-Assisted Decision Support: Clinical and Financial Outcomes. *Annals of Internal Medicine*, 124(10):884. DOI:10.7326/0003-4819-124-10-199605150-00004.

Pines, J., Selevan, J., McStay, F., George, M., and McClellan, M. (2015). Kaiser Permanente California: A Model for Integrated Care for the Ill and Injured. URL: `https://www.brookings.edu/wp-content/uploads/2016/07/KaiserFormatted\_150504RH-with-image.pdf`.

Potts, B., Adams, R., and Spadin, M. (2011). Sustaining Primary Care Practice: A Model to Calculate Disease Burden and Adjust Panel Size. *The Permanente Journal*, 15(1):53–56. PMID: 21505619.

Premier Health (2018). Premier Health. URL:`https://www.premierhealth.com/`.

Prime, L. (2005). WiPPed into shape? *Practice Nurse: The Journal for Nurses in General Practice; Philadelphia*, 30(4):7.

Public Health Agency of Canada (2001). What is the Population Health Approach? - Population Health Approach. URL:`http://www.phac-aspc.gc.ca/ph-sp/approach-approche/appr-eng.php`.

Reeves, D., Hann, M., Rick, J., Rowe, K., Small, N., Burt, J., Roland, M., Protheroe, J., Blakeman, T., Richardson, G., and others (2014). Care plans and care planning in the management of long-term conditions in the UK: a controlled prospective cohort study. *Br J Gen Pract*, 64(626):e568–e575.

Reid, R. J., Fishman, P. A., Yu, O., Ross, T. R., Tufano, J. T., Soman, M. P., and Larson, E. B. (2009). Patient-Centered Medical Home Demonstration: A Prospective, Quasi-Experimental, Before and After Evaluation. *The American Journal of Managed Care*, 15(9):e71–e87.

Reilly, J. (2013). Three key features in community healthcare changes. *British Journal of General Practice*, July. DOI: 10.3399/bjgp13X669338.

Robinson, J. H., Callister, L. C., Berry, J. A., and Dearing, K. A. (2008). Patient-centered care and adherence: Definitions and applications to improve outcomes. *Journal of the American Academy of Nurse Practitioners*, 20(12):600–607. DOI: 10.1111/j.1745-7599.2008.00360.x.

Roland, M. and Guthrie, B. (2016). Quality and Outcomes Framework: what have we learnt? *BMJ*, 354:i4060.

Sackett, D. L. (1997). Evidence-based medicine. *Seminars in Perinatology*, 21(1):3–5. DOI: 10.1016/S0146-0005(97)80013-4.

Salatino, M., De Maio, M., and Aliverti, E. (2016). *Mastering JBoss Drools 6*. Packt Publishing Ltd.

Salgado, C. M., Azevedo, C., Proena, H., and Vieira, S. M. (2016). Missing Data. In *Secondary Analysis of Electronic Health Records*, pages 143–162. Springer International Publishing. DOI: 10.1007/978-3-319-43742-2_13.

Salvatier, J., Wiecki, T. V., and Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2. DOI: 10.7717/peerj-cs.55.

Sarafidis, P. A., Lazaridis, A. A., Ruiz-Hurtado, G., and Ruilope, L. M. (2017). Blood pressure reduction in diabetes: lessons from ACCORD, SPRINT and EMPA-REG OUTCOME. *Nature Reviews Endocrinology*, 13(6):365–374. DOI: 10.1038/nrendo.2016.209.

Seddon, M., Marshall, M., Campbell, S., and Roland, M. (2001). Systematic review of studies of quality of clinical care in general practice in the UK, Australia and New Zealand. *Quality in Health Care : QHC*, 10(3):152–158. DOI:10.1136/qhc.0100152.

Segal, L. and Leach, M. J. (2011). An evidence-based health workforce model for primary and community care. *Implementation Science*, 6(93):8. DOI: 10.1186/1748-5908-6-93.

Semenov, I., Kopanitsa, G., Denisov, D., Alexandr, Y., Osenev, R., and Andreychuk, Y. (2018). Patients Decision Aid System Based on FHIR Profiles. *Journal of Medical Systems*, 42(9):166. WOS:000440484400001.

Senn, J. A. (1989). *Analysis and Design of Information Systems*. Computer Science Series. McGraw-Hill Education (UK), 2 edition. ISBN: 0-07-100606-0.

Shiffman, R. N. (1997). Representation of Clinical Practice Guidelines in Conventional and Augmented Decision Tables. *Journal of the American Medical Informatics Association*, 4(5):382–393.

Shortliffe, E. H. (1974). A Rule-based Computer Program for Advising Physicians Regarding Antimicrobial Therapy Selection. In *Proceedings of the 1974 Annual ACM Conference - Volume 2*, ACM '74, pages 739–739, New York, NY, USA. ACM. DOI: 10.1145/1408800.1408906.

Shortliffe, E. H., Davis, R., Axline, S. G., Buchanan, B. G., Green, C. C., and Cohen, S. N. (1975). Computer-based consultations in clinical therapeutics: Explanation and rule acquisition capabilities of the MYCIN system. *Computers and Biomedical Research*, 8(4):303–320. DOI:10.1016/0010-4809(75)90009-9.

SIGN (2001). The Scottish Intercollegiate Guidelines Network. URL: `http://www.sign.ac.uk/who-we-are.html`.

Silagy, C. and Weller, D. (2001). Evidence-based Practice in Primary Care: An Introduction. In Silagy, C. and Haines, A., editors, *Evidence-based Practice in Primary Care*, pages 1–11. BMJ Publisher Group, London, 2 edition. ISBN: 0-7279-1568-1.

Skipper, M. (2010). Managed clinical networks. *British Dental Journal*, 209(5):241–242. DOI: 10.1038/sj.bdj.2010.771.

Smith, S. M., Soubhi, H., Fortin, M., Hudon, C., and ODowd, T. (2012). Managing patients with multimorbidity: systematic review of interventions in primary care and community settings. *Bmj*, 345:e5205.

SNOMED International (2018). 4. SNOMED CT Basics. URL: `https://confluence.ihtsdotools.org/display/DOCSTART/4.+SNOMED+CT+Basics`.

Spackman, K. A., Campbell, K. E., and Ct, R. A. (1997). SNOMED RT: a reference terminology for health care. *Proceedings of the AMIA Annual Fall Symposium*, pages 640–644. DOI:PMC2233423.

Spiegelhalter, D. J. and Knill-Jones, R. P. (1984). Statistical and Knowledge-Based Approaches to Clinical Decision-Support Systems, with an Application in Gastroen-

terology. *Journal of the Royal Statistical Society. Series A (General)*, 147(1):35–77. DOI: 10.2307/2981737.

Starfield, B. (2001). Basic concepts in population health and health care. *Journal of Epidemiology & Community Health*, 55(7):452–454. DOI: 10.1136/jech.55.7.452.

Steinwachs, D. M. and Hughes, R. G. (2008). Health Services Research: Scope and Significance. In Hughes, R. G., editor, *Patient Safety and Quality: An Evidence-Based Handbook for Nurses*, Advances in Patient Safety. Agency for Healthcare Research and Quality (US), Rockville (MD). PMID: 21328761.

Stellefson, M. (2013). The chronic care model and diabetes management in US primary care settings: a systematic review. *Preventing chronic disease*, 10. DOI: 10.5888/pcd10.120180.

Stokes, J., Man, M.-S., Guthrie, B., Mercer, S. W., Salisbury, C., and Bower, P. (2017). The Foundations Framework for Developing and Reporting New Models of Care for Multimorbidity. *Annals of Family Medicine*, 15(6):570–577. DOI: 10.1370/afm.2150.

Struijs, J. N., Drewes, H. W., Heijink, R., and Baan, C. A. (2015). How to evaluate population management? Transforming the Care Continuum Alliance population health guide toward a broadly applicable analytical framework. *Health Policy*, 119(4):522–529. DOI: 10.1016/j.healthpol.2014.12.003.

Sweetser, A. (1999). A Comparison of System Dynamics (SD) and Discrete Event Simulation (DES). *17th International Conference of the System Dynamics Society*, pages 20–23.

Taylor, A. (2015). Building the House of Care. URL: `http://personcentredcare.health.org.uk/sites/default/files/resources/buildingthehouseofcare\_0.pdf`.

Terry, K. (2017). Wait Times for New-Patient Appointments Rise 30%. URL: `http://www.medscape.com/viewarticle/877616`.

Townsville Mackay Medicare Local (2012). *Part 2 ROLES : The Chronic Conditions Management Team in General Practice.* A Guide to Chronic Condition Management in General Practice. URL: `http://docplayer.net/17217827-Part-2-roles-the-chronic-conditions-management-team-in-gener/`

al-practice-inside-this-section-a-guide-to-chronic-condition-management.html.

Trindade, L. d. L. and Pires de, D. E. (2013). Implications of primary health care models in workloads of health professionals. *Texto & Contexto - Enfermagem*, 22(1):36–42. DOI: 10.1590/S0104-07072013000100005.

Tsiknakis, M., Katehakis, D. G., and Orphanoudakis, S. C. (2002). An open, component-based information infrastructure for integrated health information networks. *International Journal of Medical Informatics*, 68(1):3–26. DOI:1386-5056(02)00060-6.

Utley, M. and Worthington, D. (2012). Capacity Planning. pages 11–30. DOI: 10.1007/978-1-4614-1734-7_2.

Vaishnavi, V. and Kuechler, W. (2007). *Design Science Research methods and Patterns: Innovating Information and Communication Technology*. Auerbach Publications, Taylor & Francis Group, Boca Raton, FL, New York.

van Ginneken, E. and Rijken, M. (2016). ICARE4eu, a Health Programme project (2008-2013) on integrated care for multimorbidity, targeted at policymakersEwout Van Ginneken. *European Journal of Public Health*, 26(suppl_1). DOI: 10.1093/eurpub/ckw173.061.

Wagner, E. H. (1995). Population-based management of diabetes care. *Patient Education and Counseling*, 26(1-3):225–230. ISSN: 0738-3991.

Wagner, E. H. (1998). Chronic disease management: what will it take to improve care for chronic illness? *Effective clinical practice: ECP*, 1(1):2–4. ISSN: 1099-8128.

Wallace, E., Salisbury, C., Guthrie, B., Lewis, C., Fahey, T., and Smith, S. M. (2015). Managing patients with multimorbidity in primary care. *British Medical Journal*, 350:h176. DOI: 10.1136/bmj.h176.

Walley, J., Lawn, J. E., Tinker, A., de Francisco, A., Chopra, M., Rudan, I., Bhutta, Z. A., and Black, R. E. (2008). Primary health care: making Alma-Ata a reality. *The Lancet*, 372(9642):1001–1007. DOI: 10.1016/S0140-6736(08)61409-9.

Warwick, J. and Bell, G. (2007). An Evolving Systems-based Methodology for Healthcare Planning. In Kuhn, K. A., Warren, J. R., and Leong, T.-Y., editors, *Med-*

info 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems, page 23.

Weiner, M. (2009). Evidence Based Medicine. In LIU, L. and ZSU, M. T., editors, *Encyclopedia of Database Systems*, pages 1072–1073. Springer US. DOI: 10.1007/978-0-387-39940-9_156.

Weiss, K. B. (1998). Part I. A look at population-based medical care. *Disease-a-Month*, 44(8):353–369. DOI: 10.1016/S0011-5029(98)90005-0.

Weiss, S. M., Kulikowski, C. A., Amarel, S., and Safir, A. (1978). A model-based method for computer-aided medical decision-making. *Artificial Intelligence*, 11(1):145–172. DOI: 10.1016/0004-3702(78)90015-2.

WHO (1978). Declaration of Alma-Ata. Alma-Ata, USSR. WHO. URL:`http://www.who.int/publications/almaata\_declaration\_en.pdf`.

WHO (1986). WHO | The Ottawa Charter for Health Promotion. URL:http://www.who.int/healthpromotion/conferences/previous/ottawa/en/.

WHO (2002). Innovative Care for Chronic Conditions (ICCC) framework. Technical report, Geneva, World Health Organization. URL: `http://www.who.int/chp/knowledge/publications/icccglobalreport.pdf`.

WHO (2004). *ICD-10: International Statistical Classification of Diseases and Related Health Problems*, volume 3. Geneva, 10th revision edition. ISBN: 978 92 4 154654 9.

WHO (2005). Preparing a HealthCare Workforce for the 21st Centuary: The challenge of Chronic Conditions. Technical report, World Health Organisation, Geneva. URL: `http://www.who.int/chp/knowledge/publications/workforce\_report.pdf`.

WHO (2016a). *ICD -10 : International statistical classification of diseases and related health problems*. France, 5 edition. ISBN: 978 92 4 154916 5.

WHO (2016b). Integrated care models: an overview. URL: `http://www.euro.who.int/\_\_data/assets/pdf\_file/0005/322475/Integrated-care-models-overview.pdf`.

WHO (2016c). *Multimorbidity: Technical Series on Safer Primary Care.* World Health Organisation, Geneva. Licence: CC BY-NC-SA 3.0 IGO.

Wiecki, T. (2015). MCMC sampling for dummies. URL http://twiecki.github.io/blog/2015/11/10/mcmc-sampling/.

Woolf, S. H., Grol, R., Hutchinson, A., Eccles, M., and Grimshaw, J. (1999). Potential benefits, limitations, and harms of clinical guidelines. *BMJ : British Medical Journal*, 318(7182):527–530.

Young, J., Egan, T., Jaye, C., Williamson, M., Askerud, A., Radue, P., and Penese, M. (2017). Shared care requires a shared vision: communities of clinical practice in a primary care setting. *Journal of Clinical Nursing*, 26(17-18):2689–2702. DOI: 10.1111/jocn.13762.

Zhang, X. (2018). Application of discrete event simulation in health care: a systematic review. *BMC Health Services Research*, 18(1):687.

Appendix A

Appendix

A.1 Ethics Approval

Best Practice Advocacy Centre New Zealand (BPAC), a non-profit organisation that aims to promote best practice for primary health care in New Zealand, provided anonymised health data from a medical practice. Next few pages is the approval letter given by the University of Otago Human Ethics Committee (Health) for this research using the anonymised patient data provided by BPAC.

Academic Services
Manager, Academic Committees, Mr Gary Witte

HD15/016

1 July 2015

Professor S Cranefield
Department of Information Science
Division of Commerce
School of Business

Dear Professor Cranefield,

I am writing to you concerning your proposal entitled "**Workload management in Primary Care Organisations by Predicting Workload Arising from Long-term Health Conditions**", Ethics Committee reference number **HD15/016**.

The above research was submitted and reviewed as a 'Human Ethics Committee (Health) Departmental Conditional Approval of Projects using Health Information'. The outcome of that consideration was that the proposal was **approved**.

The standard conditions of approval for all human research projects reviewed and approved by the Committee are the following:

Conduct the research project strictly in accordance with the research proposal submitted and granted ethics approval, including any amendments required to be made to the proposal by the Human Research Ethics Committee.

Inform the Human Research Ethics Committee immediately of anything which may warrant review of ethics approval of the research project, including: serious or unexpected adverse effects on participants; unforeseen events that might affect continued ethical acceptability of the project; and a written report about these matters must be submitted to the Academic Committees Office by no later than the next working day after recognition of an adverse occurrence/event. Please note that in cases of adverse events an incident report should also be made to the Health and Safety Office:

http://www.otago.ac.nz/healthandsafety/index.html

Advise the Committee in writing as soon as practicable if the research project is discontinued.

Make no change to the project as approved in its entirety by the Committee, including any wording in any document approved as part of the project, without prior written approval of the Committee for any change. If you are applying for an amendment to your approved research, please email your request to the Academic Committees Office:

gary.witte@otago.ac.nz

jo.farrondediaz@otago.ac.nz

Approval is for up to three years from the date of this letter. If this project has not been completed within three years from the date of this letter, re-approval or an extension of approval must be requested. If the nature, consent, location, procedures or personnel of your approved application change, please advise me in writing.

Yours sincerely,

Mr Gary Witte
Manager, Academic Committees
Tel: 479 8256
Email: gary.witte@otago.ac.nz

c.c. Professor M Winikoff Head Department of Information Science

A.2 Drools working memory classes and attributes list

In our Drools classes and attributes, we present the initial classes and their attributes. Then we present the classes and the attribute(s) added to support our what if scenarios.

Objects used for data pre-processing and the workload prediction.

```
global java.lang.Long SimulationStartDate;
global java.lang.Long SimulationEndDate;
global java.lang.Integer RecallDecisionParameter;

/*
A Fact that denotes today for simulation
*/
declare Today
        today:long
end

/*
Patient basic details
The patientId uniquely identifies a patient.
The other features like age, gender,ethnicity is also recorded.
The regProv identifies the GP with whom the patient is registered with.
fundCode represents the funding code for the patient.
*/
declare Patient
        patientId:String
        gender:char
        age:int
        ethCode:int
        regProv:String
        fundCode:char
end

/*
Declaring the event when a Patient Visits
PatientVisit is an event that happens on the dateOfVisit and expires in 24hours.
*/
declare PatientVisit
        @role(event)
        patientId:String
        dateOfVisit:Date
end

/*
Declaring the recall as fact
Recalls are a part of the care plan of each patient.
The recalls happens within a period that is denoted as recallInFreq and actual date of visit is
determined as recallDate.
*/
```

```
declare Recall
	@role(event)

	patientId:String
	recallInFreq:int
	recallInterval: String
	recallDate:Date
end

/*
----------Declaring the Diabetic risk score as a fact ----------
Diabetic risk score denotes if the patient is at a low risk, medium, high risk or very high risk
of diabetes.
This fact is used in decision table to determine when a diabetis patient should be recalled.
*/
declare DiabeticRiskScore
	patientId:String
	riskScore:int
end

declare Script
	patientId:String
	medication_type:String
	no_of_medications:int
	dateOfPrescription:Date
end

declare CarePlan
	patientId: String
	problemCode: String
	latestReadCode: String
	carePlanCode:String
end

//----------Declaring Lab Requests, receiving the lab results and latest lab results
available ----------
//LabRequest shows when the lab requests were made for a patient.
//patientId identifies the patient whose lab tests are requested.
//labCode identifies the lab test
//labDetails gives the details about the test requested e.g BP, HbA1C and so on.
//resultDueDate is the due date when the lab results are expected.
declare LabRequest
	patientId:String
```

```
        resultDueDate:Date
        labCode:String
        labDetails:String
end

declare IndividualRecallDetail
        patientId:String
        carePlanCode:String
        recallInFreq:int
        recallInterval: String
end

/*
LatestLabResult shows that the last lab results received for a patient for a lab test.
This is maked as an event of lab results being received.
patientId identifies the patient whose lab tests are requested.
labCode identifies the lab test
labValue gives the numerical value of the results.
labTestReceived is the date when the lab results are acutally received.
*/
declare LatestLabResult
        patientId:String
        receivedDate:Date
        labresultStatus:String
        labCode:String
        labDetails:String
        labValue:double
end


/*----------------------Declaring the Classification for patient ----------
FirstClassification shows when the patient was first classified under some LTC.
patientId identifies the patient.
problemCode identifies the generic name of the LTC, e.g hypertension, diabetes and so on.
readCode gives the READCODE for the LTC which classified the patient to be having LTC.
dateOfClassification gives the date when the patient was first classified to be having LTC.
This is onset date in the Classification table.
*/
declare FirstClassification
        patientId:String
        problemCode:String
        readCode:String
        dateOfClassification:Date
end
```

/*Classification shows when the patient is classified with a LTC.
patientId identifies the patient.
problemCode identifies the generic name of the LTC, e.g hypertension, diabetes and so on.
readCode gives the latest READCODE of the LTC which classifies the patient to be having LTC now.
dateOfClassification gives the date when the patient was classified to be having LTC with the readCode.
This is CLASSDATE in the Classification table.
*/
declare Classification
	patientId:String
	problemCode:String
	latestReadCode:String
	dateOfClassification:Date
end

/*
----------CKDLABRESULT ----------
This is very similar to the lab results, but has created classify the results based on the two different lab results. eGFR and ACR are calculated based on the lab results and hence once these are calculated then, only.
*/
declare CKDLabResult
	patientId:String
	receivedDate:Date
	labresultStatus:String
	labCode:String
	labDetails:String
	labValue:**double**
end

Additional Attributes for classes to support what-if scenarios

This lists the classes and their additional attributes to support what-if scenarios presented in this work.

Patient

carePlus:boolean

Recall

recallComment:String

consulted:String

CarePlan

patientId: String

problemCode: String

latestReadCode:String

carePlanCode:String

severity:String

CareProviderRole

roleName:String

regNo:String

slotDuration:int

totalPatientsAttendedActualData:int

numberPatientsScheduledSimulation:int

dateOfAppointment:Date

canAttend:int

A.3 Read codes to Problem class mapping

In order to have information in a format required for this study we map Read codes to their respective problem class. Problem class are a more generic name of the disease identified using Read codes. For example, G20 and G27 belong to HYP_COD (Hypertension) problem class. The following pages presents the complete Drools decision table that maps Read codes to problem class(es).

RuleSet	predictionPack
Notes	This table is to assign problem code based on the simple readcode string or

RuleTable AssigningFirstClassifcationProblemCodeOnSingleReadCode

	CONDITION		ACTION	ACTION	PRIORITY	UNLOOP
	classifi : FirstClassification					
	readCode matches "$param", problemCode==null	readCode not in ($param)	classifi.problemCode = "$param";	System.out.println(classifi.patientId +" " + classifi.problemCode);System.out.println("$param"); update(classifi);		
Problem Types	**ReadCodes**	**Not ReadCode**	**Problem Code**	**Print Success**	**Salience**	**no-loop**
HyperTensionCode	**G2**		**HYP_COD**	**Problem Code hypertension is set**	**100**	**TRUE**
	G20.*					
	Gyu2					
	Gyu20					
Heart Disease	**Gyu3.***	**"Gyu31"**	**heartdisease**	**Problem Code heart disease is set**		
Stroke	**G61.***	**"G617"**	**stroke**	**Problem Code stroke is set**		
	G64.*					
	G66.*	**"G669"**				
	G6760					
	G6W					
	G6X					
	Gyu6F					
	Gyu6G					
TIADiagnosis	**ZV12D**		**TIA**	**Problem Code TIA is set**		
	Fyu55					
Diabetes	**C10.***	**"C10F8"**	**DM_COD**	**Problem Code diabetes is set**		
	C109J					
	C109K					
	C10C					
	C10D					
	C10E.*					
	C10F.*	**"C10F8"**				
	C10G.*					
	C10H.*					
	C10M.*					
	C10N.*					
	PKyP					
	66AS					
	C10P.*					
PAD	**G73**		**PAD**	**Problem Code PAD is set**		
	G73z.*	**"G73z1"**				
	Gyu74					
	G734					
	G73y					
Atrial Fabrillation	**G573.***	**"G5736"**	**Atrial**	**Problem Code Atrial is set**		
Heart Failure	**G58.***		**heartfailure**	**Problem Code heart failure is set**		
	G1yz1					
COPD	**H3**		**COPD**	**Problem Code COPD is set**		
	H31.*	**"H3101","H31y0","H3122"**				
	H32.*					
	H5832					
	H64640					
	H4641					
	Hyu30					
	Hyu31					
Asthma	**H33.***	**"H333"**	**asthma**	**Problem Code asthma is set**		
	H3120					
	H3B					
	173A					
	E10.*					
	E110.*					

Psychosis	E111.*		pyscosis	Problem Code Pyscosis is set		
	E1124					
	E1134					
	E11y.*	"E11y2"				
	E11z					
	E11zz					
	E11z0					
	E12.*					
	E13.*	"E135"				
	E2122					
	Eu2.*					
	Eu30.*					
	Eu31.*					
	Eu323					
	Eu328					
	Eu333					
	Eu32A					
	Eu329					

	RuleSet	predictionPack				
	RuleTable Assigning FirstClassifcation ProblemCode based on range of ReadCodes					
	CONDITION	CONDITION	ACTION	ACTION	ACTION	PRIORITY
	classi:FirstClassification					
	checkInRange(readCode, "$1","$2") , problemCode == null	readCode not in ($param)	classi.problemCode = "$param";	update($param);	System.out.println("$param");	
	Problem Types	Checks in range	Not ReadCode	Problem Code	classi	Print Success	Salience
"	HyperTensionCode	G24,G2z	"G24z1","G2400","G2410","G27"	HYP_COD	classi	Hypertension updated	25
	heart disease	G30B,G330z	"G310","Gyu31"	heartdisease		heartdisease updated	
		G3,G309					
		G33z,G3401					
		G342,G35X					
		G38,G3z					
	Stroke	G63y0,G63y1		stroke		stroke case updated	
		Gyu62,Gyu66					
	Heart Failure	662f,662i		heartfailure		heartfailure updated	
	COPD	H36,H3z	"H3y0","H3y1"	COPD		COPD updated	

	RuleSet	predictionPack					
	RuleTable Assigning ProblemCode						
	CONDITION	CONDITION	ACTION	ACTION	PRIORITY	UNLOOP	
	classifi : Classification						
	latestReadCode matches "$param",problemCode==null	latestReadCode not in ($param)	classifi.problemCode = "$param";	System.out.println(classifi.patientId +" " + classifi.problemCode);System.out.println("$param"); update(classifi);			
	Problem Types	ReadCodes	Not ReadCode	Problem Code	Print Success	Salience	no-loop
"	HyperTensionCode	G2		HYP_COD	Problem Code hypertension is set		
		G20.*					
		Gyu2					
		Gyu20					
	Heart Disease	Gyu3.*	"Gyu31"	heartdisease	Problem Code heart disease is set		
	Stroke	G61.*	"G617"	stroke	Problem Code stroke is set		
		G64.*					
		G66.*	"G669"				
		G6760					
		G6W					
		G6X					
		Gyu6F					

		Gyu6G					
	TIADiagnosis	ZV12D		TIA	Problem Code TIA is set		
		Fyu55					
	Diabetes	C10.*	"C10F8"	DM_COD	Problem Code diabetes is set		
		C109J					
		C109K					
		C10C					
		C10D					
		C10E.*					
		C10F.*	"C10F8"				
		C10G.*					
		C10H.*					
		C10M.*					
		C10N.*					
		PKyP					
		66AS					
		C10P.*					
	PAD	G73		PAD	Problem Code PAD is set		
		G73z.*	"G73z1"				
		Gyu74					
		G734					
		G73y					
	Atrial Fabrillation	G573.*	"G5736"	Atrial	Problem Code Atrial is set		
	Heart Failure	G58.*		heartfailure	Problem Code heart failure is set		
		G1yz1					
	COPD	H3		COPD	Problem Code COPD is set	100	TRUE
		H31.*	"H3101","H31y0","H3122"				
		H32.*					
		H5832					
		H64640					
		H4641					
		Hyu30					
		Hyu31					
a	Asthma	H33.*	"H333"	asthma	Problem Code asthma is set		
		H3120					
		H3B					
		173A					
	Psychosis	E10.*		pyscosis	Problem Code Pyscosis is set		
		E110.*					
		E111.*					
		E1124					
		E1134					
		E11y.*	"E11y2"				
		E11z					
		E11zz					
		E11z0					
		E12.*					
		E13.*	"E135"				
		E2122					
		Eu2.*					
		Eu30.*					
		Eu31.*					
		Eu323					
		Eu328					
		Eu333					
		Eu32A					
		Eu329					
	SMOKER	1371		NSMOK_COD	ProblemCode for smoker is set		
		137F		EXSMOK_COD			
		137K					
		137j					
		137l					
		137J					
		137M					
		137V					

	137h		CSMOK_COD			
	137m					
	137o					
	137	"137g"				

RuleSet predictionPack

RuleTable Assigning ProblemCode based on range of ReadCodes

	CONDITION	CONDITION	ACTION	ACTION	ACTION	PRIORITY
	classi : Classification					
	checkInRange(latestReadCode, "$1","$2") , problemCode == null	latestReadCode not in ($param)	classi.problemCode = "$param";	update(classi);	System.out.println("$param");	
Problem Types	**Checks in range**	**Not ReadCode**	**Problem Code**	**classi**	**Print Success**	**Salience**
HyperTensionCode	**G24,G2z**	**"G24z1","G2400","G2410","G27"**	**HYP_COD**	**classi**	**Hypertension updated**	**25**
heart disease	**G30B,G330z**	**"G310","Gyu31"**	**heartdisease**		**heartdisease updated**	
	G3,G309					
	G33z,G3401					
	G342,G35X					
	G38,G3z					
Stroke	**G63y0,G63y1**		**stroke**		**stroke case updated**	
	Gyu62,Gyu66					
Heart Failure	**662f,662i**		**heartfailure**		**heartfailure updated**	
COPD	**H36,H3z**	**"H3y0","H3y1"**	**COPD**		**COPD updated**	
SMOKER	**1377,137B**		**EXSMOK_COD**		**smoker status updated**	
	137N,137O					
	137S,137T					
	1372,1376		**CSMOK_COD**			
	137C,137D					
	137G,137H					
	137X,137f					

RuleSet predictionPack

RuleTable Assigning ProblemCode based on Override ReadCodes

	CONDITION	ACTION	ACTION	ACTION	PRIORITY	UNLOOP
	classi : Classification					
	latestReadCode matches "$1", problemCode ==null	classi.problemCode = "$param";	update(classi);	System.out.println("$param");		
Problem Types	**Override Code**	**Problem Code**	**classifi**	**Print Success**	**Salience**	**no-loop**
HyperTension Override Code	**21261**	**resolved**	**classi**	**Hypertension resolved**	**25**	**TRUE**
	212K					
Diabetes Override Code	**21263**			**Diabetes resolved**		
	212H					
Atrial Override Code	**212R**			**Atrial Fibrillation resolved**		
COPD Override Code	**2126F**			**COPD Resolved**		
Asthma Override Code	**21262**			**Asthma resolved**		
	212G					

A.4 Chronic Kidney Disease

In this work, we used Gfactor calculation and the ACR lab results to decide the CKD stage in a patient. The CKD stage decides the frequency of recalls for a patient. The following is the CKD care plans used in this work.

Ref	CKD Stage	ACR/PCR	Status	Clinician Review	Recall	Laboratory Tests
NICECG182.1.3.2	G1	A2		Advise: 12 monthly review using CKD module	12 months	Creatinine Electrolytes ACR
NICECG182.1.3.2	G2	A1		Advise: 12 monthly review using CKD module	12 months	Creatinine Electrolytes ACR
NICECG182.1.3.2	G2	A2		Advise: 12 monthly review using CKD module	12 months	Creatinine Electrolytes ACR
NICECG182.1.3.2	G1	A3	Stable	6 monthly review using CKD module	6 months	Creatinine Electrolytes ACR
NICECG182.1.3.2	G2	A3	Stable	6 monthly review using CKD module	6 months	Creatinine Electrolytes ACR
NICECG182.1.3.2	G3a	A1		12 monthly review using CKD module	12 months	Creatinine Electrolytes ACR
NICECG182.1.3.2	G3a	A2		12 monthly review using CKD module	12 months	Creatinine Electrolytes ACR
NICECG182.1.3.2	G3a	A3		6 monthly review using CKD module	6 months	CBC Creatinine Electrolytes ACR
NICECG182.1.3.2	G3b	A1		6 monthly review using CKD module	6 months	Creatinine Electrolytes ACR
NICECG182.1.7.8					6 months	CBC
NKFKDOQI 2002					12 months	PH Ca++ Phos
NICECG182.1.3.2	G3b	A2		6 monthly review using CKD module	6 months	Creatinine Electrolytes ACR
NICECG182.1.7.8					6 months	CBC
					12 months	PH Ca++ Phos
NICECG182.1.3.2	G3b	A3		4 monthly review using CKD module	4 months	Creatinine Electrolytes ACR
NICECG182.1.7.8					4 months	CBC
NKFKDOQI 2002					12 months	PH Ca++ Phos
NICECG182.1.3.2	G4	A1		6 monthly review using CKD module	6 months	Creatinine Electrolytes ACR
NICECG182.1.7.8					6 months	CBC
NKFKDOQI 2002					3 months	Calcium Phosphate PTH
NICECG182.1.3.2	G4	A2		6 monthly review using CKD module	6 months	CBC Creatinine Electrolytes ACR
NICECG182.1.7.8					6 months	CBC
NKFKDOQI					3 months	Calcium

Ref	CKD Stage	ACR/PCR	Status	Clinician Review	Recall	Laboratory Tests
2002						Phosphate PTH
NICECG182.1.3.2	G4	A3		4 monthly review using CKD module	4 months	Creatinine Electrolytes ACR
NICECG182.1.7.8					4 months	CBC
NKFKDOQI 2002					3 months	Calcium Phosphate PTH
NICECG182.1.3.2	G5	A1		3 monthly review using CKD module	3 months	Creatinine Electrolytes ACR
NICECG182.1.7.8					3 months	CBC
NKFKDOQI 2002					1months	Calcium Phosphate
					3 months	PTH
2013 KDIGO	G5	A2		2 monthly review using CKD module	2 monthly	Creatinine Electrolytes ACR
NICECG182.1.7.8					2 monthly	CBC
NKFKDOQI 2002					1 monthly	Calcium Phosphate
					3 monthly	PTH
NICECG182.1.3.2	G5	A3		6 weekly review using CKD module	2 monthly	Creatinine Electrolytes ACR
NICECG182.1.3.2					2 months	CBC
NKFKDOQI 2002					1 monthly	Calcium Phosphate
					3 monthly	PTH

Follow up time period in Months for Chronic Kidney Disease					
		Persistent albuminuria categories Description and Range			
		Null	A1	A2	A3
GFR categories (ml/min/ 1.73m2) Description and range	G1	No CKD	No CKD	12/12	6/12
	G2	No CKD	No CKD	12/12	6/12
	G3a	12/12	12/12	6/12	3/12
	G3b	6/12	6/12	3/12	3/12
	G4	3/12	3/12	3/12	2/12
	G5	3/12	3/12	3/12	2/12

VOL 3 | ISSUE 1 | JANUARY (1) 2013 http://www.kidney-international.org modified

http://www2.kidney.org/professionals/KDOQI/guidelines_bone/guidestate.htm

A.5 Diabetes

We followed a modified version of the diabetes care plan by the New Zealand Guidelines Group of Ministry of Health New Zealand (NZGG, 2011). The modified version of diabetes care plan used in this work as shared by the PCA is given below.

Diabetes Care Plans

Scheduling of diabetes care is dependent on the risk of complications from their diabetes. Classify the risk as: Very High, high, medium , and low. The
(Quantification of risk reflects the consensus of the Diabetes Advisory Group convened by the New Zealand Guidelines Group.)

If previous diagnosis of:
Coronary arteriosclerosis (disorder) | Cerebral arteriosclerosis | Transient ischemic attack | Atherosclerosis of arteries of the extremities
Classify as Very High Risk

Coronary Heart Disease diagnosis codes						
Include	G3... – G309. G30B. - G330z G33z. - G3401 G342. - G35X. G38.. – G3z.. Gyu3.%					53741008 Coronary arteriosclerosis (disorder)
Exclude	G310. Gyu31					
Override						
Stroke diagnosis codes						
Include	G61..% G63y0 - G63y1 G64..% G66..% G6760 G6W.. G6X.. Gyu62 – Gyu66 Gyu6F Gyu6G					65312002 Cerebral arteriosclerosis (disorder)
Exclude	G617. G669					

TIA diagnosis Codes						
Include	G65..- G654. G656.- G65zz ZV12D Fyu55					266257000 Transient ischemic attack (disorder)
Exclude						
PAD diagnostic codes						
Include	G73.. G73z.% Gyu74, G734., G73y.	XE0VP G73z. Gyu74 Xa0lV XE0VR, XaZJa				**51274000** Atherosclerosis of arteries of the extremities (disorder)
Exclude	G73z1					

Variable	Code	Criteria	Score
Hba1c mmol/mol	44TB	>55	+1
BP Systolic and diastolic mmHg	2469 246A	> 130 and > 80	+1
ACR mg/mmol	46TD	> 3	+1
eGFR CKD-EPI ml/min/1.73^2	44J3 (serum Creatinine)	<60	+1
Triglycerides and Tot Chol.	44Q 44P	>= 1.7 and >= 4.0	+1
Smoking Status*		Smoker	+1
Ethnicity	(code 20-39 or 41-43)	Maori or Pacific Island or South Asian	+1

Low risk <2
Moderate risk 2
High Risk 3

*Smoking Status is calculated by identifying the latest Smoking habit code and then analysing that code to see if the code falls into smoker, ex-smoker or never smoked.

Code Criteria	Qualifying Diagnostic Codes			Time criteria	CDS Term	CDS CODE Snomed
	Read codes v2	**~~CTV3~~**	**~~Snomed CT~~**			
Smoking habit codes						
Include	137.. - 137D. 137F. - 137H. 137J., 137K., 137M. – 137T. 137V. 137X. - 137h. 137j. 137l. 137m. 137o.	~~Ub0oo%~~	~~365981007%~~	Latest	SMOK_COD SMOK_DAT [Date of SMOK_COD (Latest)]	365980008 Finding of tobacco use and exposure (finding)
Exclude	137g.	~~XE0oo~~ ~~XaIQi%~~ ~~Ub0oq~~ ~~137L.~~ ~~XaQz~~ ~~XaXP9~~ ~~XaXP8~~ ~~XaXP6~~ ~~Ub0oo~~ ~~XaluQ~~ ~~Ub0p2~~ ~~Ub0p3~~				

Code for never smoked						
Include	1371.	~~XE0oh~~	~~266919005~~	Most recent of SMOK_COD	NSMOK_COD NSMOK_DAT	266919005 Never smoked tobacco (finding)
Codes for ex-smoker						
Include	1377. – 137B. 137F. 137K. 137N. – 137O. 137S. – 137T. 137j. 137l.	~~Ub1na%~~ ~~Ub0p1~~	~~8517006%~~ ~~228486009~~	Most recent of SMOK_COD	EXSMOK_COD	8517006 Ex-smoker
Exclude		~~XaQzw,~~ ~~XaXP8,~~ ~~XaXP6~~	~~517211000000106~~ ~~766611000000106~~ ~~766581000000100~~			
Current smoker codes						
Include	1372. – 1376. 137C. - 137D. 137G. - 137H. 137J. 137M. 137P. - 137R. 137V. 137X. - 137f. 137h. 137m. 137o. 137..	~~137R.%~~ ~~XE0og%~~ ~~137C.~~ ~~137G.~~ ~~137M.~~ ~~XaIlu~~ ~~XaItg~~ ~~XaJX2~~ ~~XaLQh~~ ~~XaWNE~~	~~77176002%~~ ~~266918002%~~ ~~160612007~~ ~~160616005~~ ~~160619003~~ ~~134406006~~ ~~44341000000100~~ ~~413173009~~ ~~203191000000107~~ ~~726831000000105~~	Most recent of SMOK_COD	CSMOK_COD CSMOK_DAT	77176002 Smoker (finding)
Exclude		~~XaXP9~~ ~~XaluQ, XE0oo~~				

Review Frequency in months

review SCTIDs	Review Name	Low Risk	Med	High	Very High
413095006	Clinical Review	6	3	3	3
26604007	HbA1c	6	3	3	3
75367002	blood pressure	6	3	3	3
271244005	lipids,	12	12	12	12
271075006	ACR	12	6	6	6
365757006	eGFR	12	3	3	3
394683006	Foot check	12	12	6	3
~~390735007~~	~~Retinal screening~~	~~24~~	~~24~~	~~24~~	~~12~~

A.6 Hypertension

Initially we started with a care plan (the following few pages) for Read coded hypertensive patients. Later we realised this practice too do not Read code all hypertensive patients. Hence, we extended the care plans for non-Read-coded patients (PCA wrote on the left hand side of hypertensive plans to record the recall frequency for non-Read-coded patients). We also added a constraint based on the anti-hypertensives taken by the LTC patients.

The following is the SQL query that differentiates patients, who are Read-coded and must be on a care plan for hypertension and are on anti-hypertensives, from patients who are non-Read-coded and must be on a care plan for hypertension based on the ant-hypertensives they are on.

Listing A.1: SQL query to create table of count of medications for patients whom the hypertension care plan will be applicable.

```
1  create table
       rx_antihypertensive_count_applied_hypertensive_plan as (
2  select idno, extract (year from rxdate) as
       last_year_of_antihypertensive, count(distinct lower(substr (
       genericname,1,6))) as count_antihypertensives
3  from scripts
4  where
5  idno||extract (year from rxdate) in (select idno||max(extract (
       year from rxdate))
6           from scripts
7           where drugcode in (select drugcode from
                antihypertensives_updated where not thergrpdesc = '
                Diuretics')
8           and idno in (select idno from
                rx_patients_problemcodes_ltc where problemcode = '
                HYP_COD')
9           and not idno in (select idno from
                rx_patients_problemcodes_ltc where problemcode in ('
                CKD','DM_COD') and not latestreadcode = 'No CKD')
10          group by idno)
11 and
12 idno||drugcode in (select idno||drugcode
13          from scripts
14          where drugcode in (select drugcode from
                antihypertensives_updated)
15          and idno in (select idno from
```

```
            rx_patients_problemcodes_ltc where problemcode = '
            HYP_COD')
16        and not idno in (select idno from
            rx_patients_problemcodes_ltc where problemcode in ('
            CKD','DM_COD') and not latestreadcode = 'No CKD'))
17 group by idno,extract (year from rxdate)
18 order by extract (year from rxdate),idno)
```

Plan of Care Hypertension

Definition of Hypertension

The care plan only applies to non diabetics and patients without CKD

i.e. G1A1 G2A1 or G1 A*null* G2A*null* or G*null* A*null* or G*null* A*any* See Current Stage CKD section in Chronic Kidney Disease SCTID 709044004

Code Criteria	Qualifying Diagnostic Codes			Time criteria
	Read codes v2	CTV3	Snomed CT	
Hypertension Codes Define if Individual identified as having Hypertension == Diagnostic Code				
Include	G2 G20% G24 - G2z Gyu2 Gyu20			
Exclude	G24z1 G2400 G2410 G27			
Codes for hypertension resolved				
Override	21261 212K			*Latest Code Date > Date of diagnostic code above*

Age	Systolic		Diastolic	Initial Review	Initial Review Duration	Ongoing review after
<80	>=180	or	>=110	1/52	6/52	3/12
<80	>=160	or	>=100	2/52	6/52	3/12
<80	>=140	or	>=90	1/12	3/12	3/12
<80	>=130	or	>=80			6/12
<80	<=100	or	<=50	1/52	6/52	3/12
<80	<=110	or	<=60	2/52	6/52	3/12
<80	<120	or	<80			3/12
>=80	>=180	or	>=110	1/52	6/52	3/12
>=80	>150	or	>90	2/52	6/52	3/12
>=80	>=140	or	>=90			3/12
>=80	>=130	or	>=80			6/12
>=80	<=100	or	<=50	1/52	6/52	3/12
>=80	<120	or	<80	2/52	6/52	3/12

Significantly Modified from: European Heart Journal (2013) 34, 2159–2219 doi:10.1093/eurheartj/eht151
Best Practice Journal 54 http://www.bpac.org.nz/BPJ/2013/August/hypertension.aspx accessed 20/10/2016

Annual

CBC

Renal Function: Sodium Potassium Creatinine

Every two years

ECG, ACR HbA1c and Lipids

A.7 LTC encodes (aka generic names) and their care plans

Table A.1 shows generic names and their corresponding care plans. There are a few exceptional cases such as CKD patients who are mapped to be on annual recall. However, the best practice followed is to recall any LTC patient in months. Therefore, we have CKD patients who are on a GENERAL_PLAN. Similarly, there are non-Read-coded hypertensive patients. Those patients are identified as NOT_HYP_COD but are on a HYP_COD_PLAN.

Table A.1: Care plans applied to generic names of LTCs used in this work.

Generic name	**Care plan name**
asthma	asthma_PLAN
Atrial	Atrial_PLAN
CKD	GENERAL_PLAN
CKD	CKD_PLAN
COPD	COPD_PLAN
DM_COD	DM_COD_PLAN
heartdisease	heartdisease_PLAN
heartfailure	heartfailure_PLAN
HYP_COD	HYP_COD_PLAN
NOT_HYP_COD	HYP_COD_PLAN
NOT_ON_PLAN	GENERAL_PLAN
PAD	PAD_PLAN
psychosis	psychosis_PLAN
stroke	stroke_PLAN

A.8 Decision table that decides individual recalls for each LTCs in a patient.

This decision table decides the recall frequency for individual patients based on the LTCs and care plans that apply to them.

RuleSet	predictionPack
No-loop	True
Notes	This decision table is for deciding the recalls and the required lab tests for each patient visits

RuleTable CKDRecallDecision

CONDITION	CONDITION	CONDITION	CONDITION		ACTION	Priority
$careplan:CarePlan			$drs: DiabeticRiskScore			
carePlanCode == "$1"	latestReadCode=="$param"	carePlanCode!="$param"	patientId==$careplan.patientId, riskScore $param	not (exists IndividualRecallDetail (patientId == $param, carePlanCode ==$careplan.carePlanCode))	System.out.println("In the decision table to insert individual recalls"); IndividualRecallDetail $recall = new IndividualRecallDetail(); $recall.recallInFreq = $1;$recall.recallInterval = "months"; $recall.patientId = $careplan.patientId; $recall.carePlanCode = $careplan.carePlanCode; insert($recall);	
Check problemCode					**decide recall**	
CKD_PLAN	**G1 A2**			$careplan.patientId	**12**	**1000**
	G1 A3				**6**	**1001**
	G2 A2				**12**	**1002**
	G2 A3				**6**	**1003**
	G3a				**12**	**1004**
	G3a A1				**12**	**1005**
	G3a A2				**6**	**1006**
	G3a A3				**3**	**1007**
	G3b				**6**	**1008**
	G3b A1				**6**	**1009**
	G3b A2				**3**	**1010**
	G3b A3				**3**	**1011**
	G4				**3**	**1012**
	G4 A1				**3**	**1013**
	G4 A2				**3**	**1014**
	G4 A3				**2**	**1015**
	G5				**3**	**1016**
	G5 A1				**3**	**1017**
	G5 A2				**3**	**1018**
	G5 A3				**2**	**1019**
DM_COD_PLAN			**<2**		**6**	**1020**
			==2		**3**	**1021**
			==3		**3**	**1022**
			>3		**3**	**1023**
COPD_PLAN					**6**	**1024**
heartdisease_PLAN					**6**	**1025**
stroke_PLAN					**6**	**1026**
TIA_PLAN					**6**	**1027**
PAD_PLAN					**6**	**1028**
Atrial_PLAN					**6**	**1029**
heartfailure_PLAN					**6**	**1030**
asthma_PLAN					**6**	**1031**
pyscosis_PLAN					**6**	**1032**
GENERAL_PLAN					**6**	**1033**

RuleSet	predictionPack
Notes	This decision table is to decide on recall for patient readcoded as hypertensive

RuleTable ReadcodedHypbpsinglevalues

CONDITION		CONDITION	CONDITION	CONDITION	ACTION	Priority
$careplan:CarePlan		$p:Patient	$labresult:LatestLabResult			
problemCode == "$1"	carePlanCode == "$1"	patientId == $careplan.patientId, age$1	labCode =="$1", (int)labValue$2, patientId == $p.patientId	not (exists IndividualRecallDetail (patientId == $careplan.patientId, carePlanCode == "$1"))	System.out.println("In the decision table to insert individual recalls"); IndividualRecallDetail $recall = new IndividualRecallDetail(); $recall.recallInFreq = $1;$recall.recallInterval = "months"; $recall.patientId = $careplan.patientId; $recall.carePlanCode = $careplan.carePlanCode; insert($recall);	
		Check problemCode			**decide recall**	
HYP_COD	**HYP_COD_PLAN**	**>0**	**2469,>=140**	**HYP_COD_PLAN**	**3**	**950**
			246A,>=90			**950**
		<80	**246A,<=60**			**950**
			2469,<=110			**950**
		>=80	**2469,<120**			**950**
			246A,<80			**950**

RuleTable Readcoded_Hyp_bp_multiple_values

CONDITION		CONDITION	CONDITION	CONDITION	ACTION	Priority
$careplan:CarePlan		$p:Patient	$labresult:LatestLabResult			
problemCode == "$1"	carePlanCode == "$1"	patientId == $careplan.patientId, age$1	labCode =="$1", (int)labValue$2 && (int)labValue$3, patientId == $p.patientId	not (exists IndividualRecallDetail (patientId == $careplan.patientId, carePlanCode == "$1"))	System.out.println("In the decision table to insert individual recalls"); IndividualRecallDetail $recall = new IndividualRecallDetail(); $recall.recallInFreq = $1;$recall.recallInterval = "months"; $recall.patientId = $careplan.patientId; $recall.carePlanCode = $careplan.carePlanCode; insert($recall);	
		Check problemCode			**decide recall**	
HYP_COD	**HYP_COD_PLAN**	**>0**	**2469,<140,>=130**	**HYP_COD_PLAN**	**6**	**940**
			246A,<90,>=80			
		<80	**246A,>60,<80**			
			2469,>110,<120			
HYP_COD	**HYP_COD_PLAN**	**>0**		**HYP_COD_PLAN**	**6**	**940**

RuleTable ReadcodedHypbpsinglevalues

CONDITION		CONDITION	CONDITION	CONDITION	ACTION	Priority
$careplan:CarePlan		$p:Patient	$labresult:LatestLabResult			
problemCode == "$1"	carePlanCode == "$1"	patientId == $careplan.patientId, age$1	labCode =="$1", (int)labValue$2, patientId == $p.patientId	not (exists IndividualRecallDetail (patientId == $careplan.patientId, carePlanCode == "$1"))	System.out.println("In the decision table to insert individual recalls"); IndividualRecallDetail $recall = new IndividualRecallDetail(); $recall.recallInFreq = $1;$recall.recallInterval = "months"; $recall.patientId = $careplan.patientId; $recall.carePlanCode = $careplan.carePlanCode; insert($recall);	
		Check problemCode			**decide recall**	
NOT_HYP_COD	**HYP_COD_PLAN**	**>0**	**2469,>=180**	**HYP_COD_PLAN**	**1**	**950**
			246A,>=110			**950**

RuleTable Readcoded_Hyp_bp_multiple_values

CONDITION		CONDITION	CONDITION	CONDITION	ACTION	Priority
$careplan:CarePlan		$p:Patient	$labresult:LatestLabResult			
problemCode == "$1"	carePlanCode == "$1"	patientId == $careplan.patientId, age$1	labCode =="$1", (int)labValue$2 && (int)labValue$3, patientId == $p.patientId	not (exists IndividualRecallDetail (patientId == $careplan.patientId, carePlanCode == "$1"))	System.out.println("In the decision table to insert individual recalls"); IndividualRecallDetail $recall = new IndividualRecallDetail(); $recall.recallInFreq = $1;$recall.recallInterval = "months"; $recall.patientId = $careplan.patientId; $recall.carePlanCode = $careplan.carePlanCode; insert($recall);	
		Check problemCode			**decide recall**	
NOT_HYP_COD	**HYP_COD_PLAN**	**<80**	**2469,<180,>=160**	**HYP_COD_PLAN**	**3**	**940**
			246A,<110,>=100			
NOT_HYP_COD	**HYP_COD_PLAN**	**>=80**	**246A,>=140,<150**	**HYP_COD_PLAN**		
			2469,>=90,<110			
NOT_HYP_COD	**HYP_COD_PLAN**	**<80**	**246A,>=140,<160**	**HYP_COD_PLAN**	**6**	**940**
			2469,>=90,<100			
NOT_HYP_COD	**HYP_COD_PLAN**	**>=80**	**246A,>150,<180**	**HYP_COD_PLAN**	**2**	
			2469,>90,<110			
NOT_HYP_COD	**HYP_COD_PLAN**	**>0**		**HYP_COD_PLAN**	**6**	**940**
NOT_HYP_COD	**HYP_COD_PLAN**			**HYP_COD_PLAN**		

A.9 Drools Rules of workload simulation

We presented the classes and their attributes used in our adaptable best practice based workload prediction model. The following presents the Drools rules that when given the patient data simulate the workload.

Rules used in our adaptable best practice based workload prediction model.

```
//// ------------------------------------SIMULATE THE REAL WORLD SCENARIO --------
--------------------
//This rule will extrapolate the l;ast visit to the first recall in the simulation
period.
rule "Update recalls that fall before Simulation startdate to a date within the
simulation period."
dialect "mvel"
salience 5000
	when
		$t:Today(today==SimulationStartDate)
		$recall:Recall(recallDate.getTime() < $t.today)
		exists IndividualRecallDetail(patientId == $recall.patientId)
		$minrecallInFreq : Number (intValue > 0) from accumulate ($ind :
IndividualRecallDetail(patientId == $recall.patientId,$recallinfreq:recallInFreq),
min($recallinfreq))
	then
		long min =  (long)$minrecallInFreq.intValue();
		Recall $newrecall = new Recall();
		$newrecall.patientId = $recall.patientId;
		$newrecall.recallInFreq = min;
		//$newrecall.problemCode = $minindrecall.problemCode;
		$newrecall.recallDate =
java.sql.Date.valueOf($recall.recallDate.toLocalDate().plusMonths(min))
		insert($newrecall);
		retract($recall);

end

rule "Recall creation during a patientVisit"
salience 250
	when
		$t:Today()
		$pv:PatientVisit(dateOfVisit.getTime() == $t.today)
		exists (IndividualRecallDetail(patientId == $pv.patientId))
		$minrecallInFreq : Number (intValue > 0) from accumulate ($ind :
IndividualRecallDetail(patientId == $pv.patientId,$recallinfreq:recallInFreq),
min($recallinfreq))

	then
		System.out.println("in rule:     Recall creation during a
patientVisit.");
		System.out.println("Minimum recall in weeks for " + $pv.patientId +
"is "+$minrecallInFreq.intValue());
		long min_freq =  (long)$minrecallInFreq.intValue();

		Recall $newrecall = new Recall();
		$newrecall.patientId = $pv.patientId;
		$newrecall.recallInFreq = min_freq;
		$newrecall.recallDate =
java.sql.Date.valueOf($pv.dateOfVisit.toLocalDate().plusMonths(min_freq))
		insert($newrecall);
		retract($pv);
end
```

```
//This rule checks if it is today and there is recall in future that falls on a
holiday then the recall is pushed to the week after.

rule "Push recalls to next week if recalls falls on a holiday"
salience 5000
	when
		$t:Today()
		$recalltoday:Recall(recallDate.getTime() > $t.today)
		not(eval (workingday($recalltoday.recallDate.getTime())))
		eval (RecallDecisionParameter == 1)
	then
		Recall $newrecall = new Recall();
		$newrecall.patientId = $recalltoday.patientId;
		$newrecall.recallInFreq = $recalltoday.recallInFreq;
		$newrecall.recallDate =
java.sql.Date.valueOf($recalltoday.recallDate.toLocalDate().plusDays(2));
		retract($recalltoday);
		insert($newrecall);
end

//This rule checks if it is today and there is recall in future that falls on a
holiday then the recall is pushed to the week before.
rule "Make recalls to previous week if recalls falls on a holiday"
salience 5000
	when
		$t:Today()
		$recalltoday:Recall(recallDate.getTime() > $t.today)
		not(eval (workingday($recalltoday.recallDate.getTime())))
		eval (RecallDecisionParameter == 2)
	then
		Recall $newrecall = new Recall();
		$newrecall.patientId = $recalltoday.patientId;
		$newrecall.recallInFreq = $recalltoday.recallInFreq;
		$newrecall.recallDate =
java.sql.Date.valueOf($recalltoday.recallDate.toLocalDate().minusDays(2));
		retract($recalltoday);
		insert($newrecall);
end

//This rule creates patient visits for a recall scheduled for today, which is a
working day.

rule "If working day today, and recalls are there, then create patient visits."
salience 5000
	when
		$t:Today()
		$recalltoday:Recall(recallDate.getTime() == $t.today)
	then
		PatientVisit $newpv = new PatientVisit();
		$newpv.patientId = $recalltoday.patientId;
		$newpv.dateOfVisit = $recalltoday.recallDate;

end
```

A.10 Preprocessing postgreSQL query

There were shortcomings in the dataset that required the data to undergo some preprocessing. The following few pages presents the PostgreSQL queries used of this preprocessing. These queries will have MedTech32 specific data.

```
alter table appointments rename to a_appointments
alter table scripts rename to a_scripts

alter table classifications rename to a_classifications;
alter table labs rename to a_labs;
alter table register rename to a_register;
alter table transactions rename to a_transactions;
alter table provider rename to a_provider;
alter table invoices rename to a_invoices;
alter table ethnicity rename to a_ethnicity;
alter table antihypertensives_updated rename to
a_antihypertensives_updated;


alter table funded_patients_register rename to
b_funded_patients_register;
--Alter table b_funded_patients_register to hold the LTC status. This
will be set to true if they are in the list of patients who have made an
LTC appointment.
Alter table b_funded_patients_register add column is_ltc boolean;
Alter table b_funded_patients_register add column is_readcoded boolean;
Alter table b_funded_patients_register add column is_on_care_plan
boolean;
Alter table b_funded_patients_register add column is_cohort boolean;

--Level 2 tables.

--C,NC and RX appointments by funded patients;
drop table if exists b_appointments_c_nc_rx_fp;
create table b_appointments_c_nc_rx_fp
as (select distinct
appt.idno,appt.whenappoint::timestamp::date,appt.prov,appt.duration,tran.
sercode
      from a_appointments appt,a_transactions tran
      where sercode in ('C','NC','RX')
      and appt.idno = tran.idno
      and appt.whenappoint::timestamp::date =
tran.whenseen::timestamp::date
      and appt.idno in (select idno from b_funded_patients_register)
)


--update the is_ltc status of patients who have an ltc appointment. n,cn
or rx appointments.
update b_funded_patients_register
set is_ltc = True
where idno in (
select distinct idno
from b_appointments_c_nc_rx_fp)


--Getting LTC related appointments.
--Get the rxdates for LTC medications
--LTC medications are those prescribed at-least three times in the
previous year OR is associated with T status for rxstatus.
--We are using first six letters of generic name and drug code to
uniquely identify a medication, esp antihypertensive.
```

```
--t_associated_scripts
drop table if exists b_T_associated_scripts;
create table b_T_associated_scripts as (
select idno,rxdate,drugcode,lower(substr (genericname,1,6))
six_genericname
from a_scripts
where rxstatus ='T'
order by idno,rxdate)


--sanity check that these T associated medications are prescribed atleast
three times in the previous year from the date of appointment. --- It
holds.
select distinct appt.idno,whenappoint,lower(substr (genericname,1,6))
from a_scripts scr, b_appointments_c_nc_rx_fp appt
where appt.idno = scr.idno
--and appt.idno = 'F1S099_B521_M042898' --and lower(substr
(genericname,1,6)) = 'beclom'
and rxdate::timestamp::date between whenappoint - interval '1 year' and
whenappoint
and appt.idno||lower(substr (genericname,1,6)) in (select
idno||lower(substr (genericname,1,6)) from a_scripts where rxstatus =
'T')
group by appt.idno,whenappoint,lower(substr (genericname,1,6))
having count(rxdate::timestamp::date)>=3

--number of active medications as on date of appointment.
--Active medications for the patient are those T-associated medications
that are prescribed within last 4 months from the date of appointment.
drop table b_number_active_ltc_medications;
create table b_number_active_ltc_medications as (
select distinct appt.idno,whenappoint::timestamp::date,count(distinct
lower(substr(genericname,1,6))) no_of_active_ltc_medications
      from b_appointments_c_nc_rx_fp appt,a_scripts src
      where src.idno = appt.idno
      and rxdate::timestamp::date between whenappoint - interval '4
months' and whenappoint
      and appt.idno||lower(substr(genericname,1,6)) in (select
idno||six_genericname  from b_T_associated_scripts)
      group by appt.idno,whenappoint
      order by appt.idno,whenappoint)

--potential ltc appointments - those c,nc and rx appointments which have
an ltc medication prescription within a three week window.
drop table b_three_window_prescription_consultation;
create table b_three_window_prescription_consultation as (
      select distinct appt.idno,whenappoint, count(distinct
lower(substr(genericname,1,6))) no_of_ltc_medications_on_appointment
      from b_appointments_c_nc_rx_fp appt, a_scripts scr
      where appt.idno = scr.idno
      and rxdate::timestamp::date between whenappoint-14 and whenappoint
+ 7
      and appt.idno||lower(substr(genericname,1,6)) in (select
idno||six_genericname  from b_T_associated_scripts)
      --and appt.idno = 'F1S099_B521_M042898'
      group by appt.idno,whenappoint
      order by idno,whenappoint
)
```

```
--a table to store values to compare count.
drop table b_med_compare;
create table b_med_compare as
(
select * from b_three_window_prescription_consultation
left join b_number_active_ltc_medications
using (idno,whenappoint)
order by idno
)

--update the compare values as zero for null values.
update b_med_compare
set no_of_active_ltc_medications = 0
where no_of_active_ltc_medications is null

--table to find those appointments as ltc appointments where there are
these many number of medications prescribed.
--if there is no active medication but a medication is prescribed
drop table b_ltc_appointment_c_nc_idno_date;
create table b_ltc_appointment_c_nc_idno_date as
(
select distinct idno,whenappoint from
b_med_compare
where (no_of_ltc_medications_on_appointment = 1 and
no_of_active_ltc_medications <=1)
or (no_of_ltc_medications_on_appointment >= 1 and
no_of_active_ltc_medications =2)
or (no_of_ltc_medications_on_appointment >= 2 and
no_of_active_ltc_medications = 3)
or (no_of_ltc_medications_on_appointment * 2 >=
no_of_active_ltc_medications and no_of_active_ltc_medications > 3)
and idno||whenappoint in (select distinct idno||whenappoint from
b_appointments_c_nc_rx_fp where sercode in ('C','NC'))
order by idno,whenappoint
)



--Get the ltc appointments including those of Rx appointments.
create table b_ltc_appointments as
(
select distinct idno,whenappoint from
b_med_compare
where (no_of_ltc_medications_on_appointment = 1 and
no_of_active_ltc_mediccations <=1)
or (no_of_ltc_medications_on_appointment >= 1 and
no_of_active_ltc_mediccations =2)
or (no_of_ltc_medications_on_appointment >= 2 and
no_of_active_ltc_mediccations = 3)
or (no_of_ltc_medications_on_appointment * 2 >=
no_of_active_ltc_mediccations and no_of_active_ltc_mediccations > 3)
and idno||whenappoint in (select distinct idno||whenappoint from
b_appointments_c_nc_rx_fp)
order by idno,whenappoint
)
```

```
--Run the below initialisation steps of Java Program.
--getCKDlabresults
--problemClassAssignmentFunction.

--update funded patients register to mark if they are readcoded.
update b_funded_patients_register
set is_readcoded = True
where idno in (select idno from prgm_patients_problemcodes_ltc)

--create table to store readcoded hypertensive patients.
create table c_hyp_readcoded_patients as
(
select distinct idno
from prgm_patients_problemcodes_ltc
where problemcode  = 'HYP_COD'
)
--readcoded hypertensive patients and count of antihypertensives for
them.
drop table c_antihypertensive_count_since_2011_readcoded_hyp;
create table c_antihypertensive_count_since_2011_readcoded_hyp as (

select idno,count( distinct lower(substr(genericname,1,6)))
from a_scripts
where drugcode in (select drugcode from a_antihypertensives_updated)
and extract (year from rxdate)>=2011
and idno in (select distinct idno from c_hyp_readcoded_patients)
group by idno
)
--nonresadcoded hypertensive patients and number of nondiuretic
antihypertensives they are on.
drop table c_antihypertensive_count_since_2011_nonreadcoded_hyp;
create table c_antihypertensive_count_since_2011_nonreadcoded_hyp as (

select idno,count( distinct lower(substr(genericname,1,6)))
from a_scripts
where drugcode in (select drugcode from a_antihypertensives_updated where
not thergrpdesc = 'Diuretics')
and extract (year from rxdate)>=2011
and not idno in (select distinct idno from c_hyp_readcoded_patients)
group by idno
)

--create a table to record the careplans applicable to the patients.
drop table c_care_plans_applicable;
create table c_care_plans_applicable(idno text,problemcode
text,latestreadcode text,careplan_code text);

--insert hyp_plan for readcoded hypertensive patients who are on
antihypertensive and not on diabetic or CKD care plan.
insert into c_care_plans_applicable(select distinct
idno,problemcode,latestreadcode, 'HYP_COD_PLAN' from
prgm_patients_problemcodes_ltc
where problemcode = 'HYP_COD'
and idno in (select idno from
c_antihypertensive_count_since_2011_readcoded_hyp)
--who are not diabetic
and idno not in (select idno from prgm_patients_problemcodes_ltc where
problemcode = 'DM_COD')
```

```
--who are not with CKD stage identified
and idno not in (select distinct idno from prgm_patients_problemcodes_ltc
where problemcode = 'CKD' and not latestreadcode = 'No CKD'));

insert into c_care_plans_applicable (
select distinct idno,'NOT_HYP_COD','NO_READCODE','HYP_COD_PLAN' from
prgm_patients_problemcodes_ltc
where not problemcode = 'HYP_COD'
and idno in (select idno from
c_antihypertensive_count_since_2011_nonreadcoded_hyp)
--who are not diabetic
and idno not in (select idno from prgm_patients_problemcodes_ltc where
problemcode = 'DM_COD')
--who are not with CKD stage identified
and idno not in (select idno from prgm_patients_problemcodes_ltc where
problemcode = 'CKD' and not latestreadcode = 'No CKD'));

--insert the respective care plans based on readcode. Those patients who
are readcoded to be HYP_COD are not inserted see query above.
--Patients with CKD and readcoded as No CKD are also not inserted.
insert into c_care_plans_applicable(select distinct
idno,problemcode,latestreadcode, problemcode||'_PLAN' from
prgm_patients_problemcodes_ltc
where not problemcode = 'HYP_COD' and not problemcode like '%SMOK%')

--update careplan as a general care plan for those latestreadcode is No
CKD.
update c_care_plans_applicable
set careplan_code = 'GENERAL_PLAN'
where latestreadcode = 'No CKD'

--insert general care plan for those who are not on any care_plan.
insert into c_care_plans_applicable(select distinct
idno,'NOT_ON_PLAN','NO_READCODE','GENERAL_PLAN'
from b_funded_patients_register
where idno not in (select idno from c_care_plans_applicable)
and is_ltc = true)

--update b_funded_patients_register with is_on_careplan to true for who
have an entry in c_care_plans_applicable.
update b_funded_patients_register
set is_on_care_plan = True
where idno in (select distinct idno from c_care_plans_applicable)


--Run java prgm to get the individual recall details.
select distinct idno from prgm_ind_recalls

update b_funded_patients_register
set is_cohort = True
where idno in (select idno from prgm_ind_recalls)

update b_funded_patients_register
set is_cohort = null
where is_ltc is null

-----Cohort identification
```

```
select count(distinct idno),careplancode
from prgm_ind_recalls
group by careplancode

select idno from (

select distinct * from prgm_ind_recalls
where idno in (select idno from prgm_ind_recalls where careplancode in
('DM_COD_PLAN','CKD_PLAN''GENERAL_PLAN'))
and not idno in (select idno from prgm_ind_recalls where not careplancode
in ('DM_COD_PLAN','GENERAL_PLAN','CKD_PLAN'))
and not idno in (select idno from prgm_ind_recalls where idno in (select
idno from prgm_ind_recalls where careplancode = 'GENERAL_PLAN')
                                    group by idno having count(distinct
careplancode)=1)
and idno in (select distinct idno from b_appointments_c_nc_rx_fp
where whenappoint between '2014-01-01' and '2014-12-31')
order by idno) as t1

select distinct idno from (

select distinct * from prgm_ind_recalls
where idno in (select idno from prgm_ind_recalls where careplancode in
('DM_COD_PLAN','GENERAL_PLAN'))
and not idno in (select idno from prgm_ind_recalls where not careplancode
in ('DM_COD_PLAN','GENERAL_PLAN'))
and not idno in (select idno from prgm_ind_recalls where idno in (select
idno from prgm_ind_recalls where careplancode = 'GENERAL_PLAN')
                                    group by idno having count(distinct
careplancode)=1)
and idno in (select distinct idno from b_appointments_c_nc_rx_fp
where extract(year from whenappoint)=2014)
order by idno) as t1


select distinct idno from (

select distinct * from prgm_ind_recalls
where idno in (select idno from prgm_ind_recalls where careplancode in
('CKD_PLAN','GENERAL_PLAN'))
and not idno in (select idno from prgm_ind_recalls where not careplancode
in ('CKD_PLAN','GENERAL_PLAN'))
and not idno in (select idno from prgm_ind_recalls where idno in (select
idno from prgm_ind_recalls where careplancode = 'GENERAL_PLAN')
                                    group by idno having count(distinct
careplancode)=1)
and idno in (select distinct idno from b_appointments_c_nc_rx_fp
where extract(year from whenappoint)=2014)
order by idno) as t1

select distinct idno from prgm_ind_recalls
where idno in (select idno from prgm_ind_recalls where careplancode in
('HYP_COD_PLAN','GENERAL_PLAN'))
and not idno in (select idno from prgm_ind_recalls where not careplancode
in ('HYP_COD_PLAN','GENERAL_PLAN'))
and not idno in (select idno from prgm_ind_recalls where idno in (select
idno from prgm_ind_recalls where careplancode = 'GENERAL_PLAN')
```

```
                                        group by idno having count(distinct
careplancode)=1)
and idno in (select distinct idno from b_appointments_c_nc_rx_fp
where extract(year from whenappoint)=2014)


select count(idno),extract (week from max(whenappoint)), extract (year
from whenappoint)
from b_ltc_appointments

select * from simulated_recalls_11_1_1
where recallinweeks > 0
order by idno

select idno from b_funded_patients_register
where idno not in (select idno from c_care_plans_applicable)
```

A.11 Mapping MedTech32 data to our dataset

A.11.1 The data mapping process

The data mapping process allows for a standardisation to occur from the underlying primary care system (PCS) to a generic primary care data set (GPCDS). The codes used within the PCS are converted to Read codes and SNOMED CT ConceptIDs (SCTID). Where there is a numeric value the numeric value is extracted from the PCS to the GPCDS. When the value contains a text field the value of the text field is used with the SCNCODE to map to a specific Read code and SCTID. The Read codes were used for the purposes of this thesis. The addition of the SNOMED CT codes has been a recent enhancement of the extraction process. For example, if a patient has an `SCNCODE` (see table 3 RDB1375) of SM and the VALUE1 Column (see table 3 RDB1376) has the text Yes then this data item is mapped to the Read code 137R which has the Term Description of Smoker. The following pagespresents tables used for this mapping.

Table 1 The mapping table used to map from the underlying GP system Measurement table.

READ_CODE_V2	SNOMED CT	SCN_CODE	COLUMN_VALUE	ISNUMERIC	VALUE_MAP_CODE	CONV_UNIT	SI_UNIT	MIN	MAX	PRECISION	SCALE	ACTIVE
2469	163030003	BP	VALUE1	TRUE	NULL	mmHg	mmHg	0	300	3	0	TRUE
246A	163031004	BP	VALUE2	TRUE	NULL	mmHg	mmHg	0	200	3	0	TRUE
137R	77176002	DIAP	VALUE16	FALSE	Yes	NULL	NULL	NULL	NULL	NULL	NULL	TRUE
137F	8517006	DIAP	VALUE16	FALSE	Past	NULL	NULL	NULL	NULL	NULL	NULL	TRUE
1371	160601007	DIAP	VALUE16	FALSE	No	NULL	NULL	NULL	NULL	NULL	NULL	TRUE
137R	77176002	SM	VALUE1	FALSE	Yes	NULL	NULL	NULL	NULL	NULL	NULL	TRUE
137F	8517006	SM	VALUE1	FALSE	Past	NULL	NULL	NULL	NULL	NULL	NULL	TRUE
1371	160601007	SM	VALUE1	FALSE	No	NULL	NULL	NULL	NULL	NULL	NULL	TRUE
137F	8517006	SM	VALUE1	FALSE	Recently quit	NULL	NULL	NULL	NULL	NULL	NULL	TRUE
137R	77176002	SMOKE	VALUE1	FALSE	Yes	NULL	NULL	NULL	NULL	NULL	NULL	TRUE
137F	8517006	SMOKE	VALUE1	FALSE	Recently quit	NULL	NULL	NULL	NULL	NULL	NULL	TRUE
1371	160601007	SMOKE	VALUE1	FALSE	No	NULL	NULL	NULL	NULL	NULL	NULL	TRUE
137F	8517006	SMOKE	VALUE1	FALSE	Past	NULL	NULL	NULL	NULL	NULL	NULL	TRUE

Table 2 The Descriptions for the columns in Table 1

Column Name	Description
READ_CODE_V2	The data items from the primary care system are mapped to Read Codes
SNOMED	The data items from the primary care system are mapped to Snomed CT ConceptIDs (SCTID)
SCN_CODE	Corresponds to the Code within the measurements table within the primary care system
COLUMN_VALUE	The column name which holds the value to be mapped to the Read and SCTID
ISNUMERIC	If the value in the column is expected to be a numeric
VALUE_MAP_CODE	When the value is text, the values used to map to the Read Code or SCTID
CONV_UNIT	The units used in the primary care system
SI_UNIT	The units used when mapped
MIN	The expected minimum numeric value
MAX	The expected maximum numeric value
PRECISION	The expected precision of the numeric value
SCALE	The expected scale of the numeric value
ACTIVE	If the mapping is currently active or not

Table 3 The column definitions for the Measurement table within the primary care system

RDB$FIELD_NAME	RDB$FIELD_NAME_1	RDB$FIELD_TYPE	RDB$FIELD_SUB_TYPE	RDB$NULL_FLAG	RDB$FIELD_LENGTH	RDB$FIELD_SCALE	RDB$CHARACTER_LENGTH
RDB$1373	PATIENTID	14	0	1	7	0	7
RDB$1374	WHENMEASURE	35	0	1	8	0	
RDB$1375	SCNCODE	37	0	1	6	0	6
RDB$1376	VALUE1	37	0		32	0	32
RDB$1377	VALUE2	37	0		32	0	32
RDB$1378	VALUE3	37	0		32	0	32
RDB$1379	VALUE4	37	0		32	0	32
RDB$1380	VALUE5	37	0		32	0	32
RDB$1381	VALUE6	37	0		32	0	32
RDB$1382	VALUE7	37	0		32	0	32
RDB$1383	VALUE8	37	0		32	0	32
RDB$1384	VALUE9	37	0		32	0	32
RDB$1385	VALUE10	37	0		32	0	32
RDB$1386	VALUE11	37	0		32	0	32
RDB$1387	VALUE12	37	0		32	0	32
RDB$1388	VALUE13	37	0		32	0	32
RDB$1389	VALUE14	37	0		32	0	32
RDB$1390	VALUE15	37	0		32	0	32
RDB$1391	VALUE16	37	0		32	0	32
RDB$1392	VALUE17	37	0		32	0	32
RDB$1393	VALUE18	37	0		32	0	32
RDB$1394	VALUE19	37	0		32	0	32
RDB$1395	VALUE20	37	0		32	0	32
RDB$1396	VALUE21	37	0		32	0	32

RDB$1397	VALUE22	37	0		32	0	32
RDB$1398	VALUE23	37	0		32	0	32
RDB$1399	VALUE24	37	0		32	0	32
RDB$1400	VALUE25	37	0		32	0	32
RDB$1401	VALUE26	37	0		32	0	32
RDB$1402	VALUE27	37	0		32	0	32
RDB$1403	VALUE28	37	0		32	0	32
RDB$1404	VALUE29	37	0		32	0	32
RDB$1405	VALUE30	37	0		32	0	32
RDB$1406	VALUE31	37	0		32	0	32
RDB$1407	VALUE32	37	0		32	0	32
RDB$1408	VALUE33	37	0		32	0	32
RDB$1409	VALUE34	37	0		32	0	32
RDB$1410	VALUE35	37	0		32	0	32
RDB$1411	VALUE36	37	0		32	0	32
RDB$1412	NOTE	37	0		128	0	128
RDB$1413	SCNOUTCOME	37	0		4	0	4
RDB$1414	SERPROVCODE	37	0	1	4	0	4
RDB$1415	SCNGROUP	37	0	1	4	0	4
RDB$1416	ROWINSERTWHEN	35	0	1	8	0	
RDB$1417	ROWINSERTSTAFF	37	0	1	4	0	4
RDB$1418	ROWINSERTCOMPUTER	37	0	1	15	0	15
RDB$1419	ROWINSERTLOCATION	14	0	1	1	0	1
RDB$1420	ROWEDITWHEN	35	0	1	8	0	
RDB$1421	ROWEDITSTAFF	37	0	1	4	0	4
RDB$1422	ROWEDITCOMPUTER	37	0	1	15	0	15
RDB$1423	ROWEDITLOCATION	14	0	1	1	0	1
RDB$1424	ROWINACTIVE	7	0	1	2	0	

A.11.2 The mapping process for laboratory results

The data mapping process allows for a standardisation to occur from the underlying primary care system (PCS) to a generic primary care data set (GPCDS). There is no agreed single coding standard for laboratory results in New Zealand. Within the PCS the returning laboratory results from the laboratory systems are stored within two tables the INBOX and the INLINE table. The PCS converts the content of the returned laboratory result blob field held in the inbox table into atomised data into the INLINE table. There are three columns that have to be used to uniquely identify the laboratory results. The constructed LABKEY from these three columns is used to map to Read codes and SNOMED CT. The Read codes were utilised for the purposes of this thesis. The addition of the SNOMED CT codes has been a recent enhancement of the extraction process. For example To map all the serum creatinine results from the PCS. There are six historic laboratory providers. One laboratory provider changed the laboratory result description (PROMPT) from CREATININE to CREATININE (SERUM). All seven combinations (Table 4) are mapped to: 44J3 Creatinine measurement, serum (procedure) 113075003

Table 4 The mapping of the laboratory test serum Creatinine

LABKEY	txtResultCode	READ_CODE_V2	SNOMED CT	MIN	MAX	PRECISION	SCALE	ISNUMERIC	ACTIVE
HTHOTAGOSQSERUM CREATININE	SQ	44J3	113075003	0.01	9999.99	7	3	1	1
LABRESULTcrCREATININE	cr	44J3	113075003	0.01	9999.99	7	3	1	1
M.P.S$CREA$CREATININE	CREA	44J3	113075003	0.01	9999.99	7	3	1	1
M.P.S$CREA$CREATININE (SERUM)	CREA	44J3	113075003	0.01	9999.99	7	3	1	1
MEDLAB94CREATININE	94	44J3	113075003	0.01	9999.99	7	3	1	1
MEDSOUTHcrCREATININE	cr	44J3	113075003	0.01	9999.99	7	3	1	1
SCLcrCREATININE	cr	44J3	113075003	0.01	9999.99	7	3	1	1

The following SQL exert was used to extract the data from the underlying PCS table

SELECT ... WHENRECEIVED AS CCIT_ENTRY_DATE, UPPER(ANYINBOX.EXTERNAPP)||'$'||RESULTCODE||'$'||UPPER(PROMPT) LABKEY, RESULTCODE, RESULT, PROMPT, ABNORM STATUS, ANYINBOX.SERPROVCODE FROM INLINE LEFT OUTER JOIN INBOX ANYINBOX ON (ANYINBOX.PATIENTID = INLINE.PATIENTID AND INLINE.WHENRECEIVED = ANYINBOX.WHENRECEIVED) LEFT OUTER JOIN PATIENT ANYBODYPAT ON (INLINE.PATIENTID = ANYBODYPAT.PATIENTID) LEFT OUTER JOIN ANYBODY ANYBODYANY ON (INLINE.PATIENTID = ANYBODYANY.ANYBODYID) WHERE INLINE.RESULT <> '' AND RESULTCODE IN ...

All the laboratory results used within the mapping /extraction process used the same methodology.

A.12 Our PyMC3 Bayesian Inference Model

The following presents implementation of our Bayesian inference model using the PyMC3 Python library.

In [1]:
```
import matplotlib.pyplot as plt
import csv
import pandas as pd
import numpy as np
import pymc3 as pm
import scipy.stats as stats
import theano.tensor as T
import random

from pandas import DataFrame
from collections import defaultdict

from pymc3 import DiscreteUniform, HalfStudentT, sample, Model, Deterministic,
 DensityDist
from pymc3 import find_MAP
from pymc3 import traceplot
from pymc3.math import minimum

from scipy.stats import binom

#Load the observed data for one patient for now.
#This data represents (startdate - visitdate)mod 90 that the patient has visit
ed in the past.
df_visit_period = pd.read_csv('C:\\My_working_data\\DataAnalysisfromDB\\Bayesi
an\\bayesian_data_june_13.csv')
df_visit_period.head()
```

Out[1]:

	idno	whenappoint	minrecallperiod	mod_no_days
0	F1S099_B521_M000000	2012-06-22	6	63
1	F1S099_B521_M000000	2013-04-24	6	9
2	F1S099_B521_M042898	2010-12-10	6	43
3	F1S099_B521_M042898	2011-05-02	6	6
4	F1S099_B521_M042898	2012-06-14	6	55

In [ ]:
```
storevalue =[]

for i in range (0, len(observed_data)):

    minrecallperiod = observed_data.index.tolist()[i][1]
    modwith = 30 * minrecallperiod
    patid = observed_data.index.tolist()[i][0]

    print "modwith = {} ".format(modwith)

    with Model() as my_model_learn_3_monthly:

        #List of (start_date - whenappoint) mod 90 for a patient.

        observed_data_for_a_patient = observed_data[i]

        alpha = HalfStudentT('alpha',lam =0.04, nu =3)

        n = 125
        medsRunOut = DiscreteUniform('medsRunOut',lower = 0, upper = (n * modw
ith)- 1)
        medsRunOutmod = Deterministic('medsRunOutmod', medsRunOut%modwith)


        def distance_between(x,y):

            x_y_minus = x-(y-modwith)
            x_y = x-y
            x_y_plus = x-(y+modwith)

            skew_factor = 5
            value1 = (skew_factor**T.sgn(x_y_minus))* (x_y_minus**2)
            value2 = (skew_factor**T.sgn(x_y))* (x_y**2)
            value3 = (skew_factor**T.sgn(x_y_plus))*(x_y_plus**2)

            min_val_1 = T.switch(T.le(value1,value2),value1,value2)
            min_val = T.switch(T.le(min_val_1,value3),min_val_1,value3)

            return min_val

        def mylogp(value):
            return_value = -alpha * distance_between(value, medsRunOutmod)
            return return_value

        PatientVisit = DensityDist('PatientVisit', mylogp, observed = observed
_data_for_a_patient)

        with my_model_learn_3_monthly:
            trace = sample(10000)

        trace_df = pm.backends.tracetab.trace_to_dataframe(trace)
        def mode(data):
            lst =[]
            hgh=0
            for i in range(len(data)):
                lst.append(data.count(data[i]))
            m = max(lst)
            ml = [x for x in data if data.count(x)==m ] #to find most frequent
 values
            mode = []
            for x in ml: #to remove duplicates of mode
                if x not in mode:
```

```
                    mode.append(x)
            print type(mode)
            return mode

        mode_mromodes = mode(pd.Series.tolist(trace_df['medsRunOutmod']))

        #if(len(mode_mromodes) > 1):
            #print "multiple modes here"
        mode_used = random.sample(mode_mromodes,1)

        mod_mro_index= trace_df.index[trace_df['medsRunOutmod'] == mode_used[0
]].tolist()

        alpha_values = trace_df.loc[mod_mro_index]

        #get the hpd values of the corresponding alphs values.

        result_alpha_hpd = pm.stats.hpd(np.asarray(alpha_values ['alpha']), al
pha=0.95)


        #get the corresponding alpha central value

        central_alpha = (result_alpha_hpd[0] + result_alpha_hpd[1])/2

        print "central_alpha = {} ".format(central_alpha)

        #append to the values as needed to write to csv file

        storevalue.append([patid,minrecallperiod,observed_data_for_a_patient,m
ode_used[0],central_alpha])
```

In [5]:
```
with open("C:\\My_working_data\\DataAnalysisfromDB\\Bayesian\\bayesian_output_
june_13.csv", "wb") as f:
    writer = csv.writer(f)
    writer.writerows(storevalue)
```

A.13 Bayesian probability calculation for each patient

For each patient, the posterior probability for medsRunOutmod and alpha is learned, nad then we chose mode of medsRunOut and alpha for each patient. We then use this medaRunOut and alpha to calculate the probability of visiting on each day in the interval period. The function below shows the function that calculates this probability.

Listing A.2: Java function that calculates the probability corresponding to each day in the recall-period interval.

```
1 private static double[] get_probability(int medsRunOut, double
    alpha, int rp) {
2         int dm = 30*rp;
3         double[] probabilities = new double[dm];
4         for(int j=0;j<(dm);j++){
5                 double log_prob = -alpha * distance_between(j,
                     medsRunOut,rp);
6                 probabilities[j]  = Math.exp(log_prob);
7         }
8         return probabilities;
9 }
10
11 private static double distance_between(int x, int mro, int rp)
    {
12                 // TODO Auto-generated method stub
13         /*
14         * If we do not use the minimum of three distances,
             there is higher probability for all the days
15         * and also if the mro happens to be say 1, then the
             probabilities towards 89 is zero, which may not be
             the case.
16         * If we do not use the minimum of three distances,
             there is higher probability for all the days
17         * and also if the mro happens to be say 1, then the
             probabilities towards 89 is zero, which may not be
             the case.
18         * We set rp =3 when we need to choose 90-day cycle and
             rp = least recall period when using their recall
```

```
            period interval to decide their next recall.
19          */
20          int modwith = 30*rp;
21          int skew_factor = 5;
22          int dist_minus = x-(mro-modwith);
23          int dist = x-mro;
24          int dist_plus = x-(mro+modwith);
25
26          int sign_minus = (int) Math.signum(dist_minus);
27          int sign = (int) Math.signum(dist);
28          int sign_plus = (int) Math.signum(dist_plus);
29
30          double dis_minus = Math.pow(skew_factor, sign_minus) *
               Math.pow(dist_minus, 2);
31          double dis = Math.pow(skew_factor, sign) * Math.pow(
               dist, 2);
32          double dis_plus = Math.pow(skew_factor, sign_plus) *
               Math.pow(dist_plus, 2);
33          double smallest = Math.min(dis_minus, Math.min(dis,
               dis_plus));
34          return smallest;
35  }
```