A location science model for the placement of POC CD4 testing devices as part of South Africa's public healthcare diagnostic service delivery model.

Louzanne Oosthuizen

Department of Industrial Engineering

University of Stellenbosch

Thesis presented in partial fulfillment of the requirements for the degree of Master of Engineering in the Faculty of Engineering at Stellenbosch University.

M.Eng (Research) Industrial

Study leader: Prof. James Bekker

March 2015

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work; that I am the sole author thereof (save to the extent explicitly otherwise stated); that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights, and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 26 November 2014

Abstract

South Africa has a severe HIV (human immunodeficiency virus) burden and the management of the disease is a priority, especially in the public healthcare sector. One element of managing the disease, is determining when to initiate an HIV positive individual onto anti-retroviral therapy (ART), a treatment that the patient will remain on for the remainder of their lifetime. For the majority of HIV positive individuals in the country, this decision is governed by the results of a CD4 (cluster of differentiation 4) test that is performed at set time intervals from the time that the patient is diagnosed with HIV until the patient is initiated onto ART. A device for CD4 measurement at the point of care (POC), the Alere PIMA™, has recently become commercially available. This has prompted a need to evaluate whether CD4 testing at the POC (i.e. at the patient serving healthcare facility) should be incorporated into the South African public healthcare sector's HIV diagnostic service provision model.

One challenge associated with the management of HIV in the country is the relatively large percentage of patients that are lost to follow-up at various points in the HIV treatment process. There is extensive evidence that testing CD4 levels at the POC (rather than in a laboratory, as is the current practice) reduces the percentage of patients that are lost to follow-up before being initiated onto ART. Therefore, though POC CD4 testing is more expensive than laboratory-based CD4 testing, the use of this technology in South Africa should be investigated for its potential to positively influence health outcomes.

In this research, a multi-objective location science model is used to generate scenarios for the provision of CD4 testing capability. For each scenario, CD4 testing provision at 3 279 ART initiation facilities is considered. For each facility, either (i) a POC device is placed at the site; or (ii) the site's

testing workload is referred to one of the 61 CD4 laboratories in the country. To develop this model, the characteristics of eight basic facility location models are compared to the attributes of the real-world problem in order to select the most suitable one for application. The selected model's objective, assumptions and inputs are adjusted in order to adequately model the real-world problem. The model is solved using the cross-entropy method for multi-objective optimisation and the results are verified using a commercial algorithm.

Nine scenarios are selected from the acquired Pareto set for detailed presentation. In addition, details on the status quo as well as a scenario where POC testing is used as widely as possible are also presented. These scenarios are selected to provide decision-makers with information on the range of options that should be considered, from no or very limited use to widespread use of POC testing. Arguably the most valuable contribution of this research is to provide an indication of the optimal trade-off points between an improved healthcare outcome due to POC CD4 testing and increased healthcare spending on POC CD4 testing in the South African public healthcare context. This research also contributes to the location science literature and the metaheuristic literature.

Opsomming

Suid-Afrika gaan gebuk onder ‘n swaar MIV- (menslike-immuniteitsgebreksvirus-)las en die bestuur van die siekte is ‘n prioriteit, veral in die openbare gesondheidsorgsektor. Een element in die bestuur van die siekte is om te bepaal wanneer ‘n MIV-positiewe individu met antiretrovirale- (ARV-)behandeling behoort te begin, waarop pasiënte dan vir die res van hul lewens bly. Vir die meeste MIV-positiewe individue in die land word hierdie besluit bepaal deur die uitslae van ‘n CD4- (cluster of differentiation 4-)toets wat met vasgestelde tussenposes uitgevoer word vandat die pasiënt met MIV gediagnoseer word totdat hy of sy met ARV-behandeling begin. ‘n Toestel vir CD4-meting by die punt van sorg (“POC”), die Alere PIMA™, is onlangs kommersieel beskikbaar gestel. Dit het ‘n behoefte laat ontstaan om te bepaal of CD4-toetsing by die POC (met ander woorde, by die gesondheidsorgfasiliteit waar die pasiënt bedien word) by die MIV-diagnostiese diensleweringsmodel van die Suid-Afrikaanse openbare gesondheidsorgsektor ingesluit behoort te word.

Een uitdaging met betrekking tot MIV-bestuur in die land is die betreklik groot persentasie pasiënte wat verlore gaan vir nasorg in die verskillende stadiums van die MIV-behandelingsproses. Heelwat bewyse dui daarop dat die toetsing van CD4-vlakke by die POC (eerder as in ‘n laboratorium, soos wat tans die praktyk is) die persentasie pasiënte wat verlore gaan vir nasorg voordat hulle met ARV-behandeling kan begin, verminder. Daarom, hoewel CD4-toetsing by die POC duurder is as toetsing in ‘n laboratorium, behoort die gebruik van hierdie tegnologie in Suid-Afrika ondersoek te word.

In hierdie studie is ‘n meerdoelige liggingswetenskapmodel gebruik om scenario’s vir die voorsiening van CD4-toetsvermoë te skep. Vir elke scenario word CD4-toetsvermoë by 3 279 ARV-inisiasie fasiliteite oorweeg. Vir elke fasiliteit word toetsvermoë verskaf deur (i) die plasing van POC-toestelle by

die fasiliteit, of (ii) verwysing vir laboratoriumgebaseerde toetsing by een van die 61 CD4-laboratoriums in die land. Die kenmerke van agt basiese fasiliteitsliggingsmodelle is met die kenmerke van die werklike probleem vergelyk om die mees geskikte model vir toepassing op die werklike probleem te bepaal. Die doelwitte, aannames en insette van die gekose model is daarna aangepas om die werklike probleem voldoende te modelleer. Die model is opgelos met behulp van die kruis-entropie-metode vir meerdoelige optimering, waarna die resultate deur middel van 'n kommersiële algoritme bevestig is.

Nege scenario's uit die verworwe Pareto-stel word uitvoerig aangebied. Daarbenewens beskryf die studieresultate die besonderhede van die status quo sowel as 'n scenario waar POC-toetsing so wyd moontlik gebruik word. Hierdie scenario's word aangebied om besluitnemers van inligting te voorsien oor die verskeidenheid moontlikhede wat oorweeg kan word, wat wissel van geen of baie beperkte tot wydverspreide gebruik van POC-toetsing. Die mees beduidende bydrae van hierdie navorsing is stellig dat dit 'n aanduiding bied van die optimale kompromie tussen 'n verbeterde gesondheidsorguitkoms weens CD4-toetsing by die POC, en verhoogde gesondheidsorgbesteding aan CD4-toetsing by die POC, in die konteks van Suid-Afrikaanse openbare gesondheidsorg. Die navorsing dra ook by tot die ligingswetenskapliteratuur sowel as tot die metaheuristiekliteratuur.

Acknowledgements

I am grateful to Professor James Bekker for his guidance and advice, the generous amounts of time he made available to me and his attention to detail.

I would like to thank the Department of Industrial Engineering at Stellenbosch University for providing me the with the necessary time and resources to complete this thesis.

I am deeply indebted to my family, Martin, Gardi and Tinus, who have encouraged and supported me in every possible way over the past three decades. Without you my life would be infinitely poorer.

Lastly, to my love, thank you for allowing me to share in your life and for sharing in mine - you have made me truly happy.

Contents

List of Figures

List of Tables

Nomenclature

Acronyms

ACILT	African Center for Integrated Laboratory Training
AIDS	Acquired immunodeficiency syndrome
ART	Anti-retroviral therapy
CD4	Cluster of differentiation 4 – the CD4 count indicates the stage of HIV or AIDS in a patient
CE	Conformité Européene a European certification denoting that a product conforms to certain safety and environmental standards
CHAI	Clinton Health Access Initiative
CSMG	Simon Fraser University's Complex Systems Modelling Group
DoH	The South African Department of Health
FDA	The Food and Drug Administration an agency within the United States Department of Health and Human Services, responsible for regulatory approval of medcial products (amongst others)
GDP	Gross domestic product
HEERO	WITS University's Health Economics and Epidemiology Research Office

HIV	Human immunodeficiency virus
LTFU	Loss to follow-up, within the context of HIV care the term refers to HIV positive individuals that no longer come to a healthcare facility to receive monitoring or treatment for the disease
MDGs	The United Nation's Millenium Development Goals
MDR-TB	Multi-drug resistant tuberculosis
MTSF	South Africa's Medium Term Strategic Framework, a document that defines the developmental framework for the country for a given electoral mandate period
NHLS	The National Health Laboratory Service
NSP	South Africa's National Strategic Plan on HIV, STDs and TB for the period 2012 – 2016
PCR	Polymerase chain reaction
PEPFAR	The US President's Emergency Plan for AIDS Relief
PMTCT	Prevention of mother-to-child transmission of HIV
POC	Point of care
QALY	Quality adjusted life years – a metric that is used in the economic analysis of different healthcare interventions
STDs	Sexually transmitted diseases
TB	Tuberculosis
UFL	The uncapacitated fixed charge location model
WHO	World Health Organisation

XDR-TB	Extensively drug resistant tuberculosis

Greek Symbols

β	The discount factor for transportation between hubs
δ	The number of patients that are initiated onto ART without having been enrolled in pre-ART care at an ART initiation site
γ_{ij}	The transportation cost per unit of demand between nodes i and j
κ	The number of patients that are initiated onto ART after having been enrolled in pre-ART care at an ART initiation site
ω	The transport cost per unit demand and per unit distance
ζ	The number of patients that would have been initiated onto ART if POC testing had been used at an ART initiation site

Roman Symbols

a_{ij}	A binary parameter indicating whether the distance between a candidate site j and a demand node i is less than the maximum acceptable service distance S
C_j	The capacity of a facility at candidate site j
D	The maximum distance between a demand node and the closest facility
d_{ij}	The distance between demand node i and a potential facility site j

E	The minimum distance between facilities
f_j	The fixed cost of locating a facility at a site j
g_{ij}	The transportation cost per unit of demand between nodes i and j
h_i	The demand at node i
M	A very large number, larger than the maximum d_{ij} value
n_i	The number of additional patients that would have been initiated on to ART at site i if POC testing had been used at that site
P	The number of facilities to locate
q_j	The cost of performing the CD4 tests assigned to the laboratory at site j
r_i	The cost of performing the POC tests assigned to ART initiation site i
S	The maximum acceptable service distance between a facility and the demand node it serves
v_i	The number of CD4 tests requested by ART initiation site i
w_j	The total CD4 testing volume assigned to the laboratory at site j
x_j	A decision variable indicating whether a facility is to be located at a candidate site j
y_{ij}	A decision variable indicating whether demand node i is served by a facility at node j
z_i	A decision variable indicating whether the demand at facility i is covered or not

Subscripts

i	Index of demand nodes
j	Index of facility sites

Terminology

10 Point Plan	The South African Department of Health's set of priorities released in response to the country's MTSF for the period 2009 to 2014
Cepheid GeneXpert device	A point of care testing device used to perform the GeneXpert MTB/RIF test
GeneXpert MTB/RIF test	A point of care test for the diagnosis of TB
Heuristic	A search procedure generally designed to solve a specific optimisation problem efficiently
Maputo Declaration	A declaration on laboratory service provision drawn up in 2008
Metaheuristic	A search procedure that may incorporate several heuristics, used for solving optimisation problems
Odds ratio	A statistical measure expressing how strongly the absence or presence of one factor is associated with the absence or presence of another factor
Pareto set	A collection of optimal solutions to a multi-objective optimisation problem
Alere PIMA™	A point of care diagnostic device for CD4 measurement
Sero-discordant couples	Couples where one partner is HIV positive and the other partner is HIV negative

Chapter 1

Introduction

In this chapter, the research problem is defined through a brief summary of the most relevant aspects of the literature study. The research objectives and methodology are described and the structure of the report is laid out.

1.1 Background

A new point-of-care (POC) device for human immunodeficiency virus (HIV) diagnostic testing, the Alere PIMA™ device for CD4 (cluster of differentiation 4) measurement, has recently become commercially available. This has prompted a need to re-evaluate the delivery of diagnostic services in the South African public healthcare sector to determine:

1. Whether this point-of-care CD4 testing device should be used; and
2. How the current diagnostic service delivery model should be adjusted to incorporate the use of this device (if it is to be used).

1.2 Problem definition

South Africa has a severe HIV burden and the management of the disease in the country is a priority, especially in the public healthcare sector. HIV affects the immune system cells, especially the CD4+ T lymphocytes. Over a period of time, HIV infection reduces CD4 levels, weakening the immune system and putting the body at risk of death due to cancer or other infections (Boyle *et al.*, 2012).

One element of managing the disease, is managing the process through which HIV positive individuals are initiated onto anti-retroviral therapy (ART). Though ART is not a cure for HIV, it slows the progression from HIV to AIDS (acquired immunodeficiency syndrome) (Boyle *et al.*, 2012) and is currently the best medical treatment available to HIV positive individuals. Once a patient has been initiated onto ART, they continue receiving this treatment for the remainder of their lifetime. Providing ART to an individual is costly and it is therefore necessary to carefully manage the initiation of patients onto this treatment.

The initiation onto ART is governed by a decision tree with two main criteria:

1. If an HIV positive person belongs to a target group (at present, target groups in South Africa include people with tuberculosis (TB) and pregnant women), this person is automatically eligible to be initiated onto ART; or

2. If an HIV positive person does not belong to one of the target groups, they are only initiated onto ART once their CD4 count falls below a set threshold.

Thus, for all HIV positive individuals that do not form part of one of the target groups, initiation onto ART is governed by the results of a CD4+ T lymphocytes cell count (henceforth referred to as a "CD4 test"). HIV positive patients whose CD4 count is above the ART initiation threshold continue receiving a CD4 test at a set interval until their CD4 cell count falls below the threshold value and they are initiated onto ART.

At present, blood samples for CD4 testing are drawn at clinics, hospitals and other healthcare facilities and collected for testing at laboratories. Once the testing is completed, the results are communicated back to the healthcare facility and then back to the patient. In this model, testing is conducted at a site that is physically removed from the healthcare facility and there is a delay between when the patient comes in to have their blood drawn for testing and when the results are available to be communicated back to the patient.

A specific challenge in the management of HIV positive patients, especially in the South African public healthcare sector, is patients that are lost to follow-up. Patients for whom a blood sample for CD4 testing was drawn at a specific healthcare facility that do not return to this healthcare facility once the results of their CD4 test are available, are an example of individuals that are deemed lost to follow-up. These individuals may have a CD4 cell count that falls below the threshold for ART initiation, yet, because

they do not return to the healthcare facility once their CD4 test results are available, they are not initiated onto ART.

One hypothesis is that eliminating the delay between when a sample is taken and when a result is available will reduce the number of patients that are lost to follow-up. This hypothesis is the key motivation for considering the use of POC diagnostic devices for CD4 testing. POC devices are placed within the healthcare facility where the patient's sample is taken, thus POC testing is different to traditonal laboratory-based testing because the testing takes place at the healthcare facility itself rather than at a remote site. POC devices are designed to conduct testing in a short time frame to enable the patient to wait at the healthcare facility for their results to become available.

At present, the Alere PIMA™ is the only commercially available POC CD4 device that is ready for implementation.

POC CD4 testing is more expensive than traditional laboratory-based CD4 testing, yet, it has been shown to have a positive impact on patient health by reducing the number of patients that are lost to follow-up before ART initiation (Wynberg *et al.*, 2014). For this reason, there is a need to evaluate different scenarios for the use of POC CD4 testing in South Africa. The scenarios that should be evaluated range from a model that relies completely on laboratory-based CD4 testing, to one that relies completely on POC CD4 testing. In addition, a range of hybrid scenarios that lie between these two extremes should also be evaluated.

Location science (an application of Operations Research) can be used to build and evaluate scenarios for the placement of this device at clinics and hospitals in the country. Scenarios can be evaluated by quantifying the likely impact (both financial and in terms of patient health) of using the Alere PIMA™ device for POC CD4 measurement at public healthcare facilities and of making adjustments to the South African diagnostic service delivery model to incorporate the use of this device. The outputs from this modelling can inform the decision-making of public healthcare managers in South Africa that are concerned with:

1. Managing public healthcare funds in such a way that the country achieves good HIV-specific healthcare outcomes (such as ensuring that a large portion of the number of eligible individuals are initiated onto ART); and

2. Managing diagnostic service delivery (including HIV-related diagnostics such as CD4 testing).

Concepts such as the current model for diagnostic service delivery in South Africa, POC testing, the ART initiation pathway, CD4 testing and the Alere PIMA™ device that were briefly referred to in this section, are discussed in detail in Chapter 2.

1.3 Aim and objectives

1.3.1 Aim

The aim of this study is to apply Operations Research to solve a location science problem related to HIV diagnostic service provision within the South African public healthcare sector.

1.3.2 Objectives

1. Investigate the status quo of (i) diagnostic CD4 testing in South Africa; and (ii) POC CD4 testing in general in order to build an understanding of the characteristics of the real-world problem that is to be solved;

2. Identify the different mathematical formulations that are used in location science;

3. Categorise the mathematical formulations that have been identified according to the characteristics of the real-world problem;

4. Determine what will be used as the mathematical basis for the formulation (and by implication whether the model will have a single or multiple objectives), and determine which factors will be taken into consideration during optimisation;

5. Mathematically formulate the real-world problem;

6. Solve the real-world problem using an appropriate method, for example an exact method (if possible), a metaheuristic or simulation; and

7. Interpret the results in terms of their implication for HIV diagnostic service provision in the South African public healthcare sector.

1.4 Research design

The research design is comprised of the following elements:

1. Unit of analysis: conceptual or non-empirical problem;
2. Meta-analytic questions (in analysing existing mathematical formulations of location science problems);
3. Interviews (in gathering expert opinions on the factors to include in the model and to set as the objective(s)); and
4. Application of Operations Research by means of a case study.

1.5 Research methodology

The research methodology followed comprised of the following main phases:

1. A literature study, which covered the following:
   (a) The current state of the SA public healthcare sector, including the expenditure on healthcare delivery and the strategic development goals for the sector;
   (b) The current model for diagnostic service delivery within the SA public healthcare sector;
   (c) The nature of POC CD4 testing as well as its likely impact on healthcare outcomes;
   (d) Possible changes to the SA public healthcare diagnostic service delivery model that can be investigated;
   (e) Industries where location science is applied;
   (f) Examples of the application of location science to healthcare; and
   (g) Existing mathematical formulations for location science.
2. Classification of existing mathematical formulations for location science according to their suitability for application to the real-world problem.

3. Case study: application of location science to an optimisation problem in the SA public healthcare diagnostic sector. The case study included:
   (a) Information gathering on the real-world problem from subject matter experts;
   (b) Input data gathering and analysis;
   (c) Mathematical formulation of the real-world problem;
   (d) Validation of the mathematical formulation using subject matter experts;
   (e) Application of a solution algorithm or metaheuristic;
   (f) Verification of the correct application of an algorithm or metaheuristic as well as of the quality of the results generated; and
   (g) Recommendations for diagnostic service delivery in the SA publich healthcare sector based on the results obtained.

1.6 Structure of the report

Chapter 2 provides a comprehensive overview of the real-world problem by (i) giving an overview of the South African healthcare sector; (ii) providing background on the diagnostic service delivery model employed in the SA public healthcare sector; (iii) describing the role of CD4 testing in the HIV treatment process; (iv) providing information on POC CD4 testing and its likely impact on healthcare outcomes; and (v) defining the possible changes to the current diagnostic service delivery model that are to be investigated in this thesis.

Chapter 3 provides a brief introduction to the field of Operations Research and the need for Operations Research in healthcare before proceeding to a more detailed discussion of location science and the types of models used in location science. The chapter concludes with examples of general applications of location science as well as healthcare-specific applications of location science. The purpose of this chapter is to motivate that Operations Research in general and location science in particular offer appropriate approaches to solving the real-world problem.

Chapter 4 is concerned with determining the most appropriate location science model for application to the real-world problem. The chapter introduces eight basic

location science models and then analyses the suitability of each of these models based on criteria such as the model objective, assumptions, inputs and outputs.

Chapter 5 presents the modelling of the real-world problem. The chapter provides detailed information on the input data analysis, the mathematical formulation as well as the solution methodology. The validation of the mathematical formulation as well as the verification of the solution methodology are also discussed in this chapter.

The results of the modelling as well as the implication for diagnostic healthcare service delivery are discussed in Chapter 6.

Chapter 7 summarises the research findings.

1.7 Conclusion: Introduction

This chapter introduced the research by summarising the real-world problem and setting out the research methodology as well as the structure of the report. Chapter 2 describes the real-world problem in more detail, starting with a description of the South African healthcare sector before providing details on diagnostic service delivery in the country.

Chapter 2

The real-world problem

The project was introduced in Chapter 1. The research problem as well as the aim and objectives of the study were defined, the research design and methodology were described and the structure of this thesis was summarised.

This chapter provides the background to the project by exploring each of the following topics:

1. The current state of the South African healthcare sector;
2. The current model of diagnostic service delivery within the South African public healthcare sector; and
3. The status quo of POC CD4 testing and its likely impact on healthcare service delivery.

2.1 The South African healthcare sector

It should be noted that there is wealth of literature available, on the state of the South African healthcare sector, healthcare spending both in South Africa and elsewhere in the world, HIV and TB in South Africa and HIV and TB in general. The aim of this chapter is not to attempt to provide a complete review of all of the literature available, but rather to give an overview in order to provide the reader with sufficient background on the context of the problem being investigated.

2.1.1 The status quo: SA healthcare sector

The South African healthcare service is divided into a public sector and a private sector. The public healthcare sector is based on the principle of universal care and the sector provides free basic healthcare for all citizens who cannot afford private healthcare (Human, 2010). A 2010 study estimated that the public sector served approximately 80% (Human, 2010) of the population, a 2009 study placed this figure above 85% (Chopra *et al.*, 2009) while in a 2013 statement the minister of health, Dr Aaron Motsoaledi, estimated this figure at 84% (South African Press Agency, 2013).

The public healthcare system is primarliy accessed through clinics and public hospitals and the standard of care varies (Human, 2010). It is widely recognised that the public healthcare system does not provide the same level of care as the private healthcare system. According to Human (2010) "reducing the great disparity between quality of care in the public and private sectors is one of South Africa's greatest challenges".

This study fits exclusively within the public healthcare domain. Though literature calls for greater integration, or at the least, greater co-operation between the public and the private sectors (Mayosi *et al.*, 2012), these two sectors are, to a large extent, operating independently of one another at this point in time.

Literature divides the development of the South African healthcare sector since 1994 into two periods. The period before 2009 is marked by the government's policies of AIDS denialism and by a lack of co-operation between the scientific and clinical communities and the government officials of the time. The period since 2009 is characterised by a new partnership between the scientific community and the government, leading to the wide-scale implementation of ART and an increased focus on bringing the TB epidemic under control (Mayosi *et al.*, 2012).

2.1.2 Performance in terms of the millenium development goals

The millenium development goals (MDGs) are a set of eight developmental goals that were formulated during the United Nations' Millenium Summit in 2000. All of the 189 states that were members of the United Nations at the time agreed to strive to achieve these goals by 2015. Achievement in terms of MDGs numbers 4, 5 and 6 can be used as a measure to assess the performance of the healthcare system:

1. MDG 4 states that infant mortality should be reduced by two thirds between 1990 and 2015. In the period between 1994 and 2008, the country's performance in terms of this goal deteriorated (Chopra *et al.*, 2009). Since 2009, the country's performance has improved (Mayosi *et al.*, 2012);

2. MDG 5 states that the number of maternal deaths per 100 000 live births should be reduced by 75% between 1990 and 2015. The country's performance with regards to this metric was essentially static for the period 1994 to 2008 (Chopra *et al.*, 2009). Data on the performance since 2009 is contradictory, however Mayosi *et al.* (2012) conclude that there might have been a slight improvement during this period; and

3. MDG 6 sets targets for combating HIV/AIDS, malaria, TB and other diseases. For both the period 1994 to 2008 (Chopra *et al.*, 2009) and the period 2009 to 2012 (Mayosi *et al.*, 2012), the Lancet's review (a series of six articles reviewing the state of the South African healthcare sector that are widely cited in literature) concludes that South Africa has made insufficient progress in terms of achieving this goal.

2.1.3 Healthcare spending

According to World Bank data (World Bank, 2013b), South African healthcare spending as a percentage of gross domestic product (GDP) is the highest in Africa over the period for which data is available (an average of 8.4% between 1995 and 2011). Figure 2.1 compares South African healthcare as a percentage of GDP with that of the rest of sub-Saharan Africa. Chopra *et al.* (2009) estimate that between 55% and 60% of healthcare spending is in the private sector. World Bank data, summarised in Figure 2.2, shows that the percentage of the South African healthcare budget spent on public healthcare has increased significantly from 39.9 % in 2006 to 47.7 % in 2011 (World Bank, 2013b).

Public healthcare spending is dominated by tertiary hospitals with thirty percent of the total public healthcare budget spent on super-tertiary hospitals located in Cape Town, Johannesburg and Durban (Chopra *et al.*, 2009). Specific programmes that target HIV/AIDS are taking up an increasing proportion of the healthcare budget. In the 2009 budget speech, the then minister of finance, Trevor Manuel, estimated that "if

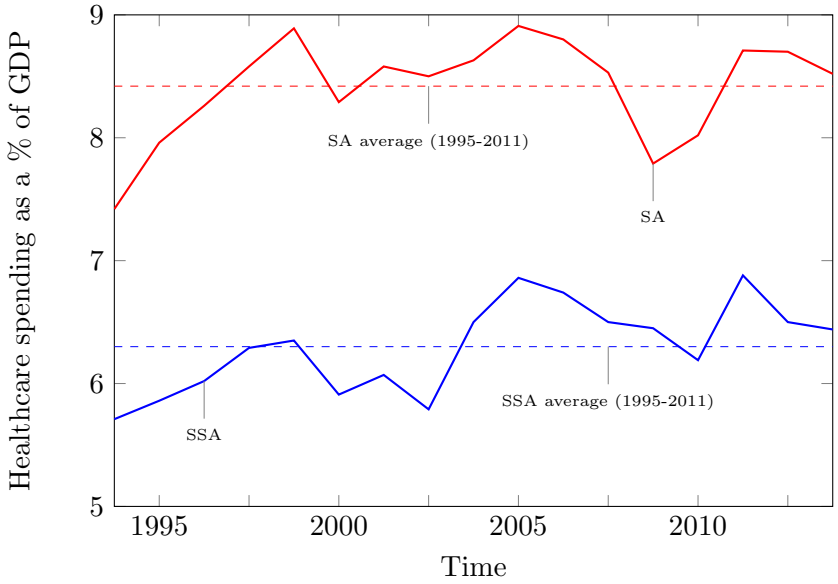

Figure 2.1: South African versus average sub-Saharan African (SSA) healthcare spending as a percentage of gross domestic product (GDP). (Data source: World Bank (2013a).)

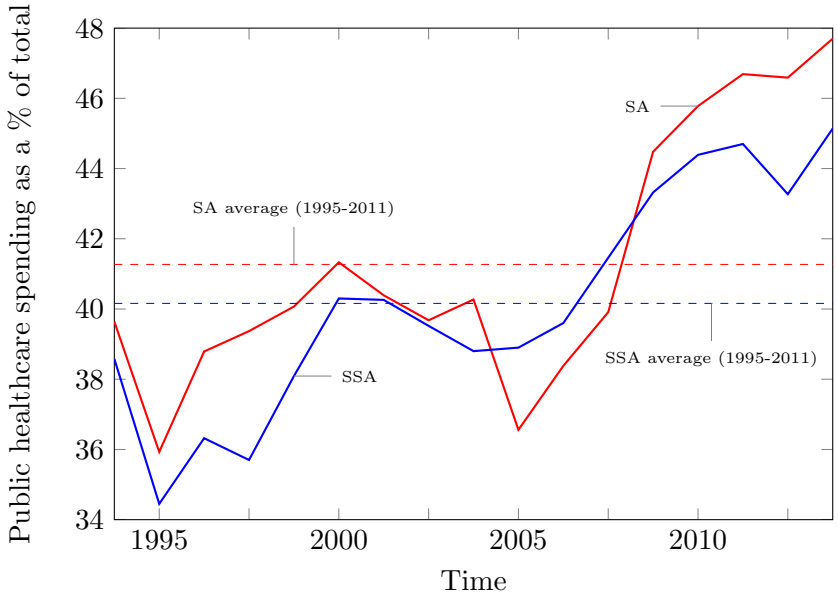

Figure 2.2: South African versus average sub-Saharan African (SSA) public healthcare spending as a percentage of total healthcare spending. (Data source: World Bank (2013c).)

the government health budget continues at its current level, 47% of it would be required to provide first-line and second-line ART (anti-retroviral therapy) for all eligible South Africans in 2014" (Mooney & Gilson, 2009).

Literature contains contradicting opinions on the sufficiency of healthcare funding. Ruff *et al.* (2011) argue that, especially in comparison to our African peers, the available healthcare funding in South Africa is adequate. The authors refer to inefficiencies in management and low productivity as the central problems that are preventing the country from extracting adequate value from the funding that is being put into the public healthcare system. In contrast to this, in an article that forms part of The Lancet's 2009 review of the South African healthcare sector Abdool Karim *et al.* (2009) argue that increased funding is necessary to rebuild physical healthcare infrastructure that has been damaged by a sustained period of neglect. The article states that "the extent to which the South African healthcare service is dysfunctional is generally underestimated" and suggests that increased spending on healthcare infrastructure would assist in adressing this problem. Several of the articles that form part of the Lancet's 2009 review describe a paradox of persistently poor health outcomes in spite of the relatively high healthcare spending in the country (Chopra *et al.*, 2009). In its 2010/2011 – 2012/2013 framework document, the Department of Health (DoH) appears to agree with the Lancet review's observations when it states that the country's healthcare outcomes are not always commensurate with its GDP ranking (Department of Health, 2010).

2.1.4 Strategic goals and objectives

South Africa's Medium Term Strategic Framework (MTSF) for the electoral mandate period 2009 – 2014 was formally accepted by Cabinet in January 2010 (Department of Health, 2010). This framework set the high-level developmental strategy for the country as a whole and was intended to guide planning and the allocation of resources in all government departments (The Presidency: Republic of South Africa, 2009). The MTSF was drawn up in line with the country's long-term goal of achieving the MDGs (The Presidency: Republic of South Africa, 2009). The MTSF contains 20 key outcomes which are to be achieved by the healthcare sector (Department of Health, 2010):

1. Increased life expectancy at birth;

2. Reduced child mortality
3. Decreased maternal mortality ratio;
4. Managing HIV prevalence;
5. Reduced HIV incidence;
6. Expanded PMTCT programme;
7. Improved TB case finding;
8. Improved TB outcomes;
9. Improved access to ART for HIV-TB co-infected patients;
10. Decreased prevalence of multi-drug-resistant tuberculosis;
11. Revitalisation of Primary Health Care (PHC);
12. Improved physical infrastructure for healthcare delivery;
13. Improved patient care and satisfaction;
14. Accreditation of health facilities for quality;
15. Enhanced operational management of health facilities;
16. Improved access to human resources for health;
17. Improved healthcare financing;
18. Strengthened Health Information Systems (HIS);
19. Improved health services for the youth; and
20. Expanded access to home based care and community health workers.

In response to the MTSF, the DoH released a set of priorities commonly referred to as the "10 Point Plan" (Health Systems Trust, 2010). These priorities were drawn up in line with the country's MTSF and aim to assist the country in achieving the MDGs (Department of Health, 2010):

1. Provision of strategic leadership and creation of a social contract for better health outcomes;
2. Implementation of a National Health Insurance (NHI) plan;
3. Improving quality of health services;
4. Overhauling the healthcare system and improving its management;
5. Improving human resources planning, development and management;
6. Revitalisation of physical infrastructure;
7. Accelerated implementation of the HIV and AIDS and sexually transmitted infections National Strategic Plan 2007 – 2011 and increased focus on TB and other communicable diseases;
8. Mass mobilisation for better health for the population;
9. Review of the drug policy; and
10. Strengthening research and development.

Appendix A contains a table that describes the key activities to be executed to achieve each of the priorities in the 10 Point Plan and indicates how these priorities and activities align to the twenty healthcare outcomes listed in the MTSF. The appendix also contains details on the quantified 2009 performance as well as the 2014 target for each of the twenty MTSF healthcare outcomes.

The National Strategic Plan (NSP) on HIV, sexually transmitted diseases (STDs) and TB 2012 – 2016 was drawn up by the South African National AIDS Council. The NSP is intended to determine the strategic direction for the national, provincial, district and community-level response to HIV, STDs and TB and it is aligned to the MTSF (South African National AIDS Council, 2012). The NSP sets out the following four strategic objectives:

1. Addressing social and structural barriers to HIV, STD and TB prevention, care and impact;
2. Preventing new HIV, STD and TB infections;

3. Sustaining health and wellness; and

4. Increasing the protection of human rights and improving access to justice.

2.1.5 HIV / AIDS in South Africa

South Africa has an estimated 6.1 million HIV positive individuals living in the country (UNAIDS, 2013). This accounts for approximately 17.9% of the estimated 34 million HIV positive individuals in the world (UNAIDS, 2011). There is consensus in literature that this is the largest number of HIV positive individuals living in any one country in the world. The country's HIV epidemic has stabilised in recent years with data indicating that the spread of the disease is starting to plateau (South African National AIDS Council, 2012). The number of new infections per year decreased with 22% between 2001 and 2009 (UNAIDS, 2011). Despite this, the country still has the largest number of new infections per annum in the world (UNAIDS, 2011).

South Africa has the largest number of individuals that are co-infected with HIV and TB in the world (UNAIDS, 2011) and TB is the leading cause of death for HIV positive individuals in the country (UNAIDS, 2013). UNAIDS (2013) estimates that 330 000 of the 520 000 new TB cases diagnosed in South Africa in 2011 were co-infected individuals. TB was declared a national emergency in the country in 2004 and South Africa currently has the third largest TB burden in the world (South African National AIDS Council, 2012). The high multi-drug resistant tuberculosis (MDR-TB) and extensively drug resistant tuberculosis (XDR-TB) caseload has further complicated efforts to combat the disease (UNAIDS, 2013). South Africa has adopted an integrated strategy for the management of TB and HIV aimed at preventing HIV positive individuals from developing TB (UNAIDS, 2013). An estimated 31% of co-infected individuals were receiving treatment for both TB and HIV and an estimated 102 000 co-infected individuals were receiving ART in 2012 (UNAIDS, 2013). The number of deaths attributed to TB in co-infected individuals in the country has decreased from 99 000 in 2004 to 88 000 in 2012 (UNAIDS, 2013).

There is no known cure for HIV, however, worldwide, ART is the preferred method of treatment for people with HIV. HIV positive individuals who receive ART have a lower likelihood of becoming ill or dying from AIDS, developing TB, and transmitting both HIV and TB (UNAIDS, 2013). The positive impact of ART on health is illustrated

by the 11.3 year increase in life expectancy in KwaZulu-Natal between 2003 (when ART scale-up began) and 2011 (UNAIDS, 2013).

South Africa has the largest ART programme in the world. An estimated 1.8 million people were receiving ART in April 2011 (Mayosi *et al.*, 2012), by December 2012 the minister of health estimated that the figure had increased to 1.9 million people (Motsoaledi, 2012) while UNAIDS (2013) estimated the 2012 figure at 2.01 million people. To place the scale of the South African ART programme in perspective, India has the second largest ART programme in the world with an estimated 570 000 individuals receiving treatment (UNAIDS, 2013). According to the World Health Organisation (WHO) 2010 guidelines, 81% of the approximately 2.5 million people who were eligible for ART in 2012 were receiving the treatment. The WHO published a new set of guidelines for ART eligibility in 2013, under these guidelines, the number of eligible individuals in the country increases to 5.1 million people (UNAIDS, 2013). At the time of writing this thesis, South Africa was not complying with these new guidelines. It is evident that it would require significant financial resources to essentially double the number of individuals receiving ART. Table 2.1 lists the 20 biggest ART programmes in the world and states the estimated ART coverage based on WHO 2010 guidelines.

The annual number of AIDS-related deaths in the country has reduced by 21% between 2001 and 2010 (UNAIDS, 2011). In the 2011 World AIDS day report, UNAIDS (2011) predicted that the substantial increase in the number of people receiving ART between 2009 and 2010 in South Africa would most likely result in a significant reduction in the number of new infections in the country. The organisation estimated that, at the current levels of healthcare spending, the number of new infections will stabilise at approximately 500 000 per year (UNAIDS, 2011).

Table 2.1: Size and estimated coverage of the twenty biggest ART programmes in the world (Data source: UNAIDS (2014)).

Country	**People on ART** [1]	**People needing ART** [2]	**Estimated ART coverage** [2,3]
South Africa	2 150 881	2 695 092	80%
India	604 987	NA	50%
Kenya	604 027	827 624	73%
Zimbabwe	565 675	716 375	79%
Nigeria	491 021	1 537 092	32%
Zambia	480 925	605 416	79%
Uganda	438 542	681 587	64%
United Republic of Tanzania	432 293	710 095	61%
Malawi	405 131	584 394	69%
Brazil	313 175	NA	NA
Mozambique	309 851	688 620	45%
Ethiopia	288 137	478 875	61%
Thailand	239 090	NA	76%
Botswana	212 083	211 946	95%
China	126 448	NA	NA
Russian Federation	125 623	NA	NA
Cameroon	122 783	275 662	45%
Namibia	116 687	129 221	90%
Rwanda	114 618	132 045	87%
Côte d'Ivoire	110 370	226 509	49%

[1] Based on Global AIDS Response Progress Reporting data.

[2] Based on national Spectrum files.

[3] Based on WHO 2010 Guidelines.

Abbreviations: NA, not available.

2.2 Diagnostic service delivery in South Africa

Accurate diagnostic testing is essential for clinical diagnosis, monitoring of infectuous diseases and directing public healthcare policy (Petti *et al.*, 2006). In developed countries, the majority of clinical decisions are based on the outcomes of diagnostic laboratory tests and the appropriate use of laboratory medicine is increasingly being recognised as a strategy for assisting countries in achieving the MDGs (Elbireer *et al.*, 2011). Nkengasong (2010) states that improving clinical and public health laboratories is a key component of health systems strengthening.

2.2.1 The current service delivery model

Diagnostic service provision in the country exists in both the private sector and the public sector. Within the public healthcare sector, diagnostic testing capabilities are provided exclusively by the National Health Laboratory Service (NHLS) (National Health Laboratory Service Website, 2014). In addition to routine and specialist diagnostic pathology testing facilities in all of the nine provinces, the NHLS contains four specialist units: (i) the National Institute for Communicable Diseases; (ii) the National Institute for Occupational Health; (iii) the National Cancer Registry; and (iv) the South African Vaccine Producers (National Health Laboratory Service Website, 2014).

The NHLS employs a tiered system of service provision where small primary and district laboratories offer basic testing capabilities and selected sites in larger metropolitan areas and academic hospitals offer more specialist testing capabilities (in addition to the basic testing capabilities). This tiered system is aligned to the Maputo Declaration on laboratory service provision that recommends that a tiered, integrated laboratory network provides the best model for diagnostic service provision in resoure-limited settings (Maputo Conference, 2008). The Maputo Declaration recommends four levels of testing facilities:

1. Level I – Primary laboratories;
2. Level II – District laboratories;
3. Level III – Regional or Provincial laboratories; and
4. Level IV – National or Reference laboratories.

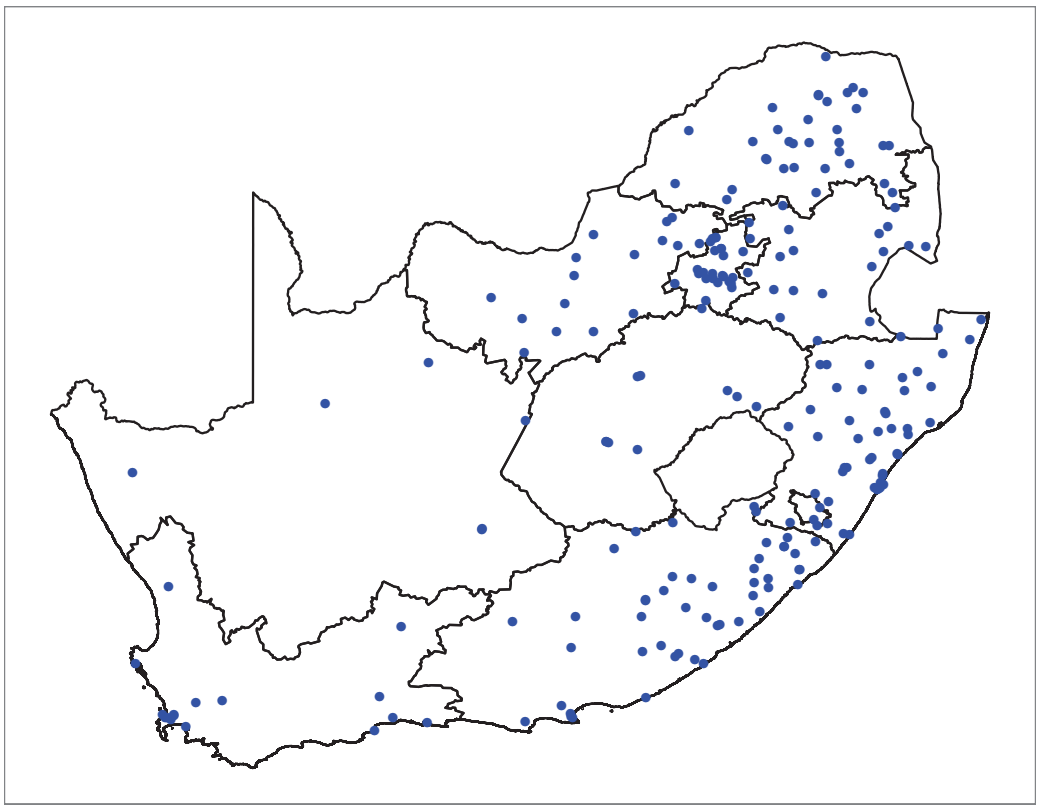

Figure 2.3: The distribution of NHLS laboratories.

The NHLS operates in excess of 300 laboratories in the country, according to the criteria defined in the Maputo Declaration, eight of these can be classified as Level IV National or Reference laboratories (National Health Laboratory Service Website, 2014). A 2006 study of diagnostic testing capablity in sub-Saharan Africa makes a generalised statement that access to laboratory testing facilities is severly limited, but moderates this statement by remarking that capabilities of laboratories vary widely accross the region (Petti *et al.*, 2006). Though access to laboratory testing in many sub-Saharan African countries may be severly limited, it is clear that this cannot be said to apply to South Africa. However, though the number of laboratories in the country may be sufficient, it is less clear whether the equity of distribution of laboratory facilities is satisfactory. When viewing the geographical distribution of laboratory facilities depicted in Figure 2.3, these appear to be densely situated in some parts of the country and sparsely situated in others.

The NHLS relies on a well-developed transport network, generally operated by couriers and porters that are employed by the NHLS, to transport diagnostic specimens from clinics and hospital wards to local laboratories. All specimens are received and registered on the NHLS Laboratory Information System at the local laboratory before being referred on to other laboratories for more specialised testing, as required. The NHLS also offers private pathology providers in the country a referral service for tests that are not requested frequently or are expensive to provide.

The South African Minister of Health was one of seven African Ministers of Health that signed the African Society for Laboratory Medicine's 2012 Ministerial Call for Action to Strengthen Laboratory Services in Africa (International Conference of the African Society for Laboratory Medicine, 2012). The Call for Action acknowledges the Maputo Declaration, it places emphasis on the importance of ensuring that there is a sufficiently large, professional, quality laboratory workforce and declares a commitment towards developing harmonised policies to govern the quality of diagnostic products and medical devices in Africa (International Conference of the African Society for Laboratory Medicine, 2012).

South Africa also provides support for laboratory services elsewhere in Africa. An example of this is the South African National Accreditation System that is used to accredit certain laboratories outside of the country's borders (Olmsted *et al.*, 2010). Another example is the NHLS' National Institute for Communicable Diseases that, in collaboration with the WHO and PEPFAR (President's Emergency Plan for AIDS Relief), established the African Center for Integrated Laboratory Training (ACILT) (Nkengasong *et al.*, 2010). ACILT provides training to laboratory personnel from across Africa and ACILT training staff perform follow-up visits to provide additional support (Nkengasong *et al.*, 2010).

2.2.2 Point of care testing

Ehrmeyer & Laessig (2007) summarised the many definitions of point of care (POC) testing as follows: "patient specimens assayed at or near the patient with the assumption that test results will be available instantly or in a very short timeframe to assist caregivers with immediate diagnosis and / or clinical intervention". POC is a relatively new development in diagnostic testing that shifts the point of analysis from a laboratory environment, that is typically situated at some distance from the patient, to the

point of care (i.e. to the patient's bedside or to the clinic consultation room). In the South African context, POC testing devices have also been situated at primary and district laboratories as a means of decentralising analytical capabilities and bringing these closer to the patient. The characteristic of POC testing are discussed in more detail in the next subsection.

The following POC tests are relevant within the management of the TB and HIV epidemics in South Africa:

1. The Cepheid GeneXpert MTB/RIF test for the diagnosis of TB. This test is currently being used as part of South Africa's diagnostic service delivery model. It is discussed in more detail in Subsection 2.2.2.2;

2. Rapid / simple tests for the diagnosis of HIV. These tests are also currently being used as part of the country's diagnostic service delivery model and are discussed in more detail in Subsection 2.2.2.3;

3. CD4 measurement for the measurement of HIV. A POC CD4 test has recently become commercially available, it is the purpose of this study to recommend whether this device should be used in South Africa and, if it is to be used, how it should be integrated with the existing diagnostic service delivery model. This is introduced in Subsection 2.2.2.4 and discussed in more detail in Section 2.3; and

4. HIV Viral load monitoring. Viral load monitoring forms part of the management process for patients that are on ART. POC viral load testing technology is not currently available, although it is expected that such technology will become available in the future. Ths is referred to briefly in Subsection 2.2.2.4.

2.2.2.1 Charateristics of POC

POC testing varies widely in terms of the complexity of the analysis. For example, many rapid HIV tests rely solely on the reaction that occurs when a drop of blood is placed on a test strip. In contrast to this, the GeneXpert device that is used widely throughout South Africa for TB analysis, involves a full polymerase chain reaction (PCR) analysis that takes place within the bench-top analyser. In spite of the varying complexity of the analysis, a key characteristic of a POC test is that the operator

should require no or minimal scientific training in order to execute the test accurately (Anderson *et al.*, 2011).

Another key characteristic is that the reagents and consumables required for the test need to be able to withstand a wide range of temperatures and have a long shelf life, this sets POC apart from typical laboratory tests where the monitoring and control of reagents and their storage conditions is a necessary component of quality control. Anderson *et al.* (2011) suggest a shelf life exceeding at least six months at ambient conditions (where ambient conditions can vary from more than 40°C to below 0°C). The ability of the test itself to be performed outside of a temperature-controlled environment is also important (Anderson *et al.*, 2011).

There is consensus in literature that POC testing should be "rapid". There is no exact definition for the time-frame that constitutes rapid testing, and POC testing time varies from a few minutes (e.g. in the case of HIV diagnosis) to a few hours (e.g. in the case of GeneXpert MTB/RIF for TB). However, there is consensus that the testing should occur while the patient waits, so that, even in cases where it is not possible to initiate treatment immediately, the patient leaves the clinical encounter with certainty regarding their diagnosis and with a clear idea of the way forward (Pai *et al.*, 2012).

2.2.2.2 Existing POC testing in South Africa: TB diagnosis

Pai *et al.* (2012) describes POC testing as a particularly powerful tool for the management and control of infectious diseases. In South Africa, rapid testing for HIV diagnosis and GeneXpert MTB/RIF for TB diagnosis have been implemented on a large scale in recent years.

In 2010, the WHO endorsed the use of the GeneXpert MTB/RIF test for patients who are suspected of having either MDR-TB or being co-infected with HIV and TB, this description encompasses the majority of suspected TB patients in South Africa (Meyer-Rath *et al.*, 2012). At this point in time, the most common method of testing for TB was through a combination of smear microscopy and culture. The smear microscopy technique is more than 125 years old and is unable to detect drug-resistance (Van Rie *et al.*, 2010), it also has a poor ability to detect TB in HIV positive individuals[1]. Culture for the detection of TB is the so-called "gold standard" of testing, however the test has

[1] Getahun *et al.* (2007) cites 15 different publications to give a range of between 24% and 61% of HIV positive TB patients testing smear negative.

a turnaround time of between two and six weeks and requires a biosafety level[1] two or three facility, which cannot be easily provided in remote settings (Van Rie *et al.*, 2010). Long turnaround time is a particular concern in the diagnosis of MDR-TB, with a study conducted prior to the implementation of Cepheid GeneXpert technology finding that 33% of MDR-TB patients passed away and 16% could not be traced by the time that the laboratory-confirmed MDR-TB diagnosis (based on culture) was available (Heller *et al.*, 2010).

In March 2011, the DoH announced a nationwide implementation of the Cepheid GeneXpert device to replace smear microscopy. The goal was to improve the accuracy of TB detection and to ensure that drug resistant TB cases are detected early so that appropriate treatment can be initiated timeously (Mayosi *et al.*, 2012). The implementation was to be achieved in a two to three year timeframe (Meyer-Rath *et al.*, 2012).

As of 2012, South Africa has the largest Cepheid GeneXpert MTB/RIF testing programme in the world (Mayosi *et al.*, 2012) with 37% of global Cepheid GeneXpert instrument sales and 53% of global Cepheid GeneXpert cartridge sales being attributed to the country (Health Systems Trust, 2013). By March 2013, South Africa had more than 290 Cepheid GeneXpert devices being operated in more than 140 centres and almost 1.2 million tests had been performed (UNAIDS, 2013).

In a study that aimed to quantify the likely impact and cost of the Cepheid GeneXpert implementation in South Africa, Meyer-Rath *et al.* (2012) estimated that, once the implementation of the device had been completed, an additional 30% to 37% of TB cases would be diagnosed per year (this would include an estimated additional 69% to 71% of MDR-TB cases). Furthermore, the study found that 81% of patients would be diagnosed after their first visit (versus 46% before the Cepheid GeneXpert implementation). The study estimated that the cost of TB diagnosis alone would increase by 55% while the total treatment cost per TB case would increase by 8% (Meyer-Rath *et al.*, 2012). In South Africa's 2013 budget estimate, the National Treasury credits the

[1] A facility's biosafety level classification is determined based on a wide range of factors, including (i) attributes of the physical infrastructure such as the ventilation system, access control and the availability of emergency infrastructure such as eye wash stations; (ii) the level of training of the personnel as well as the type of supervision that is present in the facility; and (iii) the standard operating procedures that govern safety aspects such as the handling of sharp materials, the transport of specimens and materials, and the decontamination of work surfaces.

implementation of the Cepheid GeneXpert technology with strengthening the country's TB subprogramme and allocates an additional R338 million to cover the higher testing costs associated with the Cepheid GeneXpert (National Treasury, 2013).

At present, the Cepheid GeneXpert is being implemented in laboratories (including those located at district hospitals) throughout the country. Lessells *et al.* (2013) describes the design of a study to compare the impact on the timely initiation of appropriate TB treatment when the Cepheid GeneXpert is located in the primary healthcare clinic itself rather than centrally at a district hospital. This study is currently under way.

2.2.2.3 Existing POC testing in South Africa: HIV diagnosis

POC testing for the diagnosis of HIV is widely used in South Africa and around the world. In 2010, the South African DoH introduced a voluntary HIV counselling and testing initiative with the aim of expanding its ART programme. Twelve million people were tested during a twelve month period (Glencross *et al.*, 2012).

There are a large number of companies that produce and supply assays for the detection of HIV antibody. The WHO states that of the 29 assays that have been evaluated in recent years, 19 were so-called "simple" or "rapid" tests that do not require any equipment other than the consumables and reagents used (World Health Organisation, 2014). In its latest report on the operational characteristics of HIV rapid diagnostic tests, the World Health Organisation (2013b) evaluated eight assays, with a testing time that varied between two minutes and 17 minutes for a single specimen. Plate (2007) reviewed the implementation of POC for HIV diagnosis in 11 African countries, as part of the review 15 rapid assays were evaluated. The study defines rapid HIV tests as tests that provide results within 30 minutes and require minimal equipment and training. Eight of the 15 assays that were evaluated were based on the lateral flow principle, these assays do not require any reagents other than that contained in the test device itself and none of these test devices require refrigeration. Wu & Zaman (2012) agree that rapid tests that are based on the lateral flow principle are the most ready-to-use, the study also defines rapid tests as assays that provide results within 30 minutes and require minimal or no reagents or equipment. Due to the high sensitivity and high specificity achieved by rapid HIV assays, these tests are included alongside

other laboratory-based tests for HIV diagnosis as gold standard technologies (Wu & Zaman, 2012).

In a review of low-cost diagnostics tools, Wu & Zaman (2012) credit rapid HIV tests with significantly increasing both (i) the number of individuals that are tested for HIV; and (ii) the number of HIV positive individuals that receive their results before being lost to follow-up. Anderson *et al.* (2011) describe the positive effect that rapid POC tests for HIV diagnosis has had on healthcare provision, especially in remote areas where timely availability of HIV results have: (i) assisted in the prevention of mother-to-child transmission (PMTCT); (ii) improved the uptake of voluntary testing; and (iii) reduced the loss of patients to follow-up.

2.2.2.4 The next step in POC testing: CD4 measurement

As discussed in Subsection 1.1, a device for POC CD4 testing, the Alere PIMA™ has recently become commercially available. Evidence suggests that the use of POC CD4 testing for governing the initiation onto ART can improve healthcare outcomes by reducing the number of patients that are lost to follow-up (Wynberg *et al.*, 2014). Consequently, there is a need to evaluate whether this device should be used as part of diagnostic service provision in South Africa's public healthcare sector. This is the real-world problem being investigated in this thesis. This real-world problem is discussed in detail in the following section.

2.3 The real-world problem: Should POC CD4 testing be implemented in South Africa

Section 1.2 briefly described the ART initiation pathway, the role of CD4 testing in this process and the concept of loss to follow-up. These concepts are now discussed in more detail.

2.3.1 The role of CD4 testing in the ART initiation pathway

CD4 testing is used as the mechanism that governs ART initiation for the majority of HIV positive patients in South Africa. The exception is target groups that are initiated onto ART as soon as HIV is diagnosed. Target groups include pregnant women, patients with TB and sero-discordant couples .

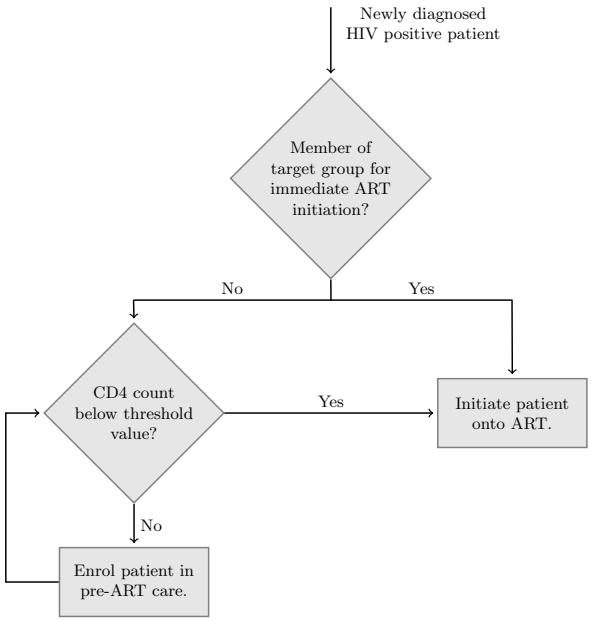

Figure 2.4: ART initiation decision tree.

Figure 2.4 depicts the process from HIV diagnosis to the initiation of ART. As shown, patients that do not form part of target groups for immediate ART initiation have their CD4 count tested after they are diagnosed as being HIV positive. If their CD4 count falls below the threshold value, they are initiated onto ART. If their CD4 cell count falls above the threshold value, their CD4 cell count is tested at a set time interval until it falls below the threshold value and they are initiated onto ART. Currently, WHO regulations stipulate an ART initiation threshold of 500 cells/mm^3 (World Health Organisation, 2013a). As stated in Subsection 2.1.5, South Africa is still operating according to the previous WHO recommended threshold of 350 cells/mm^3. In South Africa, the set interval for CD4 testing prior to ART initiation is six months for individuals with a CD4 count of less than 500 cells/mm^3 and twelve months for individuals with a CD4 count of more than 500 cells/mm^3 (Lessells *et al.*, 2011).

Though it is not necessary for determining their eligibility for ART, patients that form part of the target groups for immediate ART initiation also receive a CD4 test following a positive HIV diagnosis. Thus, any person that tests HIV positive, should have a CD4 test performed following this diagnosis.

After ART initiation, patients receive both CD4 tests and HIV viral load tests at

set intervals as part of the monitoring process for patients on ART. At present, HIV viral load testing is not available as a POC test, as such, it falls outside the scope of this study. CD4 testing as part of the monitoring process for patients receiving ART also falls outside the scope of this discussion, as it is not expected that implementing POC at this point will have any impact on loss to follow-up, the primary driver of improved health outcomes being investigated in this study. However, if POC CD4 testing capability is provided at a site, it is reasonable to assume that healthcare workers at this site will use the POC device for all CD4 tests performed at this site. As such, the volumes of CD4 tests that are performed as part of ART monitoring at each site must be included in the scope of this study.

According to the Maputo declaration, CD4 testing should be offered at all Level II, Level III and Level IV laboratories (Maputo Conference, 2008).

2.3.2 Loss to follow-up and the expected impact of CD4 testing

The concept of loss to follow-up was briefly described in Section 1.2 and referred to again in Subsections 2.2.2.3 and 2.2.2.4. Within the context of HIV care, the term "loss to follow-up", is used to refer to HIV positive patients that no longer come to a healthcare facility to receive monitoring or treatment for this disease. There are a number of points in the HIV care pathway where patients can be lost to follow-up. The term "lost to programme" includes both patients that have been lost to follow-up as well as patients that have passed away during any given stage of the care pathway.

2.3.2.1 Loss to follow-up in the HIV care pathway

Mugglin *et al.* (2012) conducted a systematic review of 29 studies from sub-Saharan Africa in order to estimate the losses at various stages prior to initiation onto ART. Figure 2.5 depicts the pathway from HIV diagnosis to initiation onto ART and indicates where losses occur.

The left section of Figure 2.5 shows the stage from HIV diagnosis to CD4 testing. As mentioned in Subsection 2.3.1 any patient that has tested positive for HIV (e.g. through a rapid test for HIV diagnosis) should have a CD4 test performed, regardless of whether this patient forms part of one of the target groups for ART initiation. However, in a meta-analysis of studies covering the period from HIV diagnosis to ART initiation, Mugglin *et al.* (2012) estimate that only 72% of patients that test positive

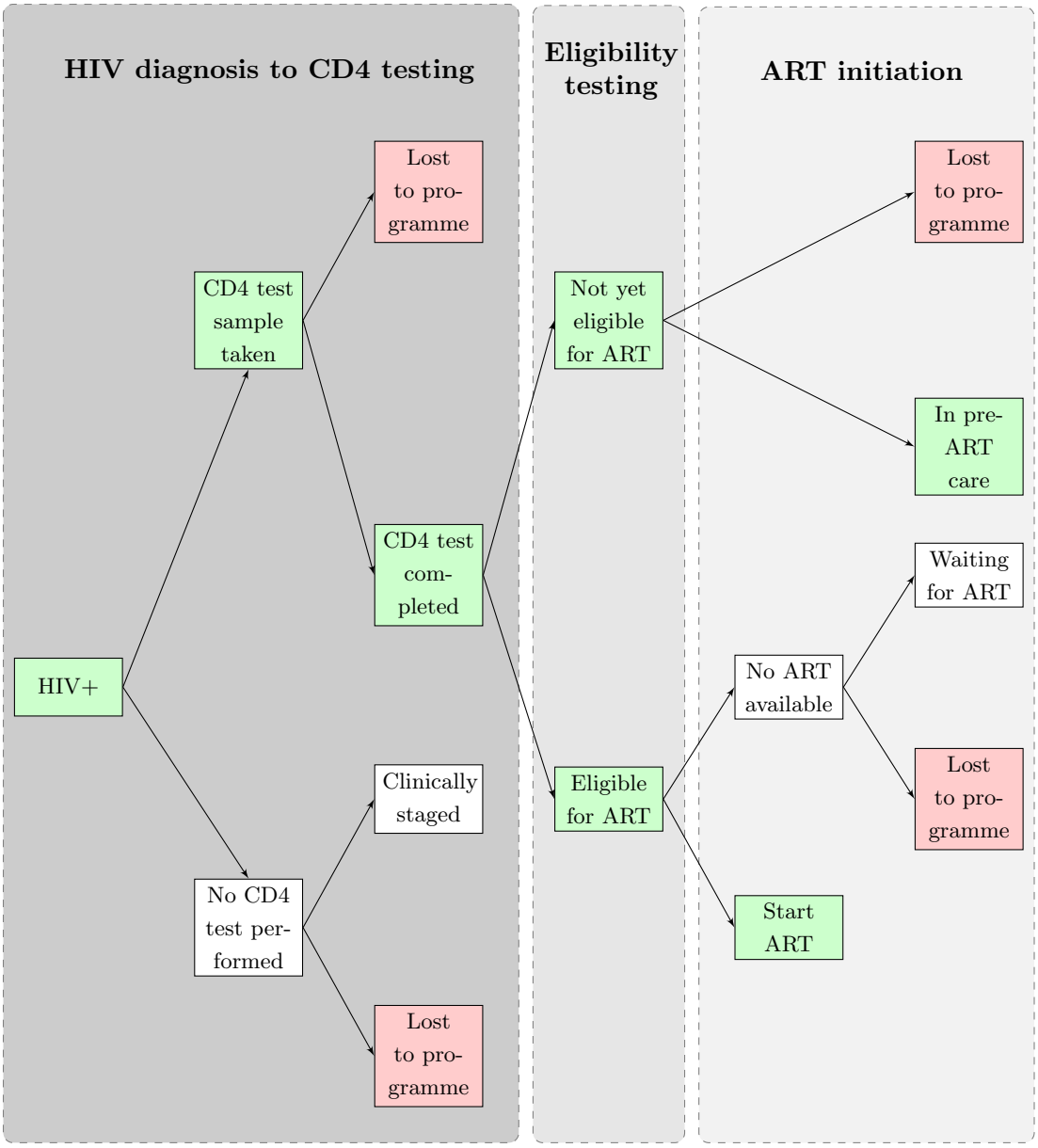

Figure 2.5: Routes from HIV diagnosis to ART initiation (Mugglin *et al.*, 2012).

for HIV have their CD4 count measured ("CD4 test sample taken" in Figure 2.5). The figure shows that a portion of HIV positive patients that do not receive a CD4 cell count are clinically staged, this clinical staging is not a part of the standardised pre-ART care pathway and it is unclear whether these patients eventually start ART. The remainder of the patients who do not receive a CD4 test at this point are considered lost to the programme. (As stated in Subsection 2.3.2 "Lost to programme" includes patients that are lost to follow-up as well as patients that have passed away.)

The study classified a CD4 count as being completed ("CD4 test completed" in Figure 2.5) if the CD4 count was performed within a set time frame after HIV diagnosis and the patient was informed of the CD4 result. Few of the studies reviewed by Mugglin *et al.* (2012) measured the percentage of patients that received the result of their CD4 test. In a study conducted at a large HIV/AIDS clinic in South Africa between January 2008 and February 2009, Larson *et al.* (2010) found that 65% of patients who had a sample taken for CD4 testing did not complete this testing within 12 weeks of testing positive for HIV.

The middle section of Figure 2.5 depicts stage two, assessment of eligibility for ART. Through meta-analyses, Mugglin *et al.* (2012) found that 40% of patients that tested positive for HIV were eligible for ART initiation at this stage.

The right section of Figure 2.5 depicts stage three, from eligibility assessment to the start of ART. As shown in the figure and discussed previously in this section, patients that are not yet eligible for ART should be enrolled into pre-ART care (where they will continue receiving a CD4 test at set time intervals until they are eligible for ART initiation). Patients that do not enter pre-ART care are considered lost to the programme. Though it falls outside the bounds of the Mugglin *et al.* (2012) study, it is important to note that patients that are enrolled in pre-ART care may stop coming for CD4 tests and become lost to follow-up prior to ART initiation.

Through meta-analysis, Mugglin *et al.* (2012) found that 63% of patients that are eligible for ART are successfully initiated onto treatment. As shown in Figure 2.5, the remainder of these patients are either waiting for ART or are lost to the programme. The "waiting for ART" step is not part of the official pre-ART care pathway although studies show that some of these patients are eventually initiated onto ART.

Though it falls outside of both the scope of this thesis and the scope of the study conducted by Mugglin *et al.* (2012), it is important to note that not all of the patients

that are initiated onto ART continue with this treatment until death, these patients are categorised as lost to follow-up after initiation of ART.

In summary, Mugglin *et al.* (2012) found that in total 13% of ART eligible patients and 57% of ART ineligible patients were lost to follow-up before the initiation of ART. As mentioned in Subsection 2.3.2.1, the 57% does not include the number of patients that were enrolled in pre-ART care but were later lost to follow-up before being initiated onto ART.

2.3.2.2 The expected impact of POC CD4 testing on loss to follow-up

In recent years, several publications (including Anderson *et al.* (2011), Zachariah *et al.* (2011) and Boyle *et al.* (2012)) have argued that POC CD4 testing has the potential to reduce the loss to follow-up in the HIV care pathway. POC CD4 testing could be conducted immediately after rapid HIV diagnosis so that a newly diagnosed patient would know on the same day of treatment whether they are eligible for ART. With the exception of patients who choose not to have a CD4 test conducted after testing positive for HIV, such an approach could potentially ensure that all newly diagnosed HIV positive patients have their ART eligibility determined after diagnosis (the end of stage two in Figure 2.5). Even in cases where CD4 testing is not conducted on the same day as HIV diagnosis, POC CD4 testing should eliminate the loss to follow-up between having a CD4 test conducted and receiving the CD4 result. (This loss was estimated by Larson *et al.* (2010) and is discussed in Subsection 2.3.2.1.) Finally, it is reasonable to assume that POC CD4 testing would have an impact on the loss to follow-up rate of patients enrolled in pre-ART care.

Zachariah *et al.* (2011) give the following reasons for an urgent need for POC CD4 testing in low income contries:

1. To assess eligibility for ART;
2. To start ART earlier;
3. To improve PMTCT uptake (this is not applicable in South Africa as pregnant women form part of the target group that are initiated onto ART regardless of CD4 count);
4. To simplify ART (initiation) at the primary care level;

5. To enhance task shifting (i.e. to ensure that ART initiation can be conducted at remote healthcare facilities). This has limited applicability to South Africa where ART initiation is already conducted at remote healthcare facilities; and

6. To reduce early attrition from programmes (this refers to the loss to follow-up prior to ART initiation).

Following the commercial availability of POC CD4 testing capabilities, a number of studies have measured the impact of POC CD4 testing on loss to follow-up. Wynberg *et al.* (2014) conducted a systematic review of 15 studies (published between 2011 and 2013) and found that:

1. POC CD4 testing increased the percentage of HIV positive patients that proceed to have a CD4 test;

2. POC CD4 testing increased the percentage of patients that received the result of their CD4 test;

3. POC CD4 testing reduced the time between HIV diagnosis and CD4 testing;

4. POC CD4 testing reduced the time from CD4 testing to receiving the test result; and

5. There is mixed evidence for increased ART initiation due to POC CD4 with two South African studies finding that patients were respectively three times more likely and six times more likely to initiate and a third South African study finding that there was no difference in the likelihood of eligible patients initiating ART.

The expected impact of POC CD4 testing on the pre-ART initiation loss to follow-up rate is quantified in Subsection 5.1.8.

2.3.3 CD4 testing methods

This section begins with a description of laboratory-based CD4 testing technology before providing an overview of the emerging field of POC CD4 testing technology.

2.3.3.1 Laboratory-based CD4 testing

CD4 testing is traditionally laboratory based. The current gold standard technology for CD4 measurement is flow cytometry (Wu & Zaman, 2012). By 2008, the simplified single platform PanLeucogated CD4 methodology was being used in 52 NHLS laboratories in South Africa (Glencross *et al.*, 2008) and it is the predicate methodology used by the NHLS (Glencross *et al.*, 2012). This methodology makes use of flow cytometry devices such as the BD FACSCalibur™ and the BD FACSCount™ to perform CD4 testing. This method forms the reference standard against which POC CD4 testing in South Africa should be evaluated.

2.3.3.2 POC CD4 testing technology

Both Anderson *et al.* (2011) and Zachariah *et al.* (2011) stated that reliable, cost-effective, and time-effective POC CD4 testing was not yet widely available. Anderson *et al.* (2011) made specific mention of the limited commercial release of the Alere PIMA™ analyser and its disposable test cartridges as a start to meeting the need for POC CD4 testing technology. A year later, Wu & Zaman (2012) conducted a review of mechanisms for diagnosing and monitoring HIV in low-resource settings and found that a number of technologies, including the Alere PIMA™ had recently become available. In the same year, Boyle *et al.* (2012) conducted a review of emerging technologies for CD4 testing and identified nine existing or emerging POC CD4 test technologies, including the Alere PIMA™. Table 2.2 is taken from Boyle *et al.* (2012) and summarises the identified POC CD4 technologies and their specifications.

As shown, the PointCareNOW™, CyFlow® miniPOC and Alere PIMA™ are the only POC CD4 technologies that are being marketed for which the instrument has already been developed. Of these three existing commercially available POC CD4 testing methods, the Alere PIMA™ is the only one that does not make use of flow cytometry (Boyle *et al.*, 2012). The PointCareNOW™ requires specialist skills (phlebotomy) to take a venous sample in order to provide a sufficient amount of blood for the test. It is essential that adequate quality control measures are in place to ensure that specimen preparation, equipment and reagents are functional, especially in POC settings where testing is performed outside of the controls that are associated with a traditional laboratory environment. Boyle *et al.* (2012) mention only the Alere PIMA™ as incorporating

Table 2.2: Specifications of POC CD4 technologies that are either in development or on the market. (Reproduced from: Boyle *et al.* (2012).)

Instrument	Sample volume (l)	Stabilised reagents	CD4 / CD4%	Time to result (min)	Instrument required	Automation / throughput (per hour)	Data storage	Battery run / time per charge (h)	Marketed	Quality control	Regulatory status	Dimensions: LxWxH (cm^3)[1]	Mass (kg)[1]
PointCare NOW™	40	Y	Y / Y	8	Y	Y / 5 – 6	Y	Y / 4	Y	Y	FDA / CE	33.7 x 25.0 x 35.4	<12.00
CyFlow® miniPOC	20	Y	Y / Y	1	Y	Y / 30 – 40	Y	Y / 4 – 5	Y	Y	CE	18.6 x 26.8 x 24.3	<5.00
Alere PIMA™	25	Y	Y / N	20	Y	Y / 30 – 40	Y	Y / 8	Y	Y	CE	22.0 x 13.0 x 16.0; 9.7 x 3.2 x 0.6	2.54; 0.11
Coulter® CD4 Count Kit	100	N	Y / N	30	Y	N / 2	N	N	Y	Y	CE	NA	NA
Dynal T4 Quant Kit	125	N	Y / N	30	Y	N / 6	N	N	Y	NA	CE	NA	NA
Daktari	10	Y	Y / N	8	Y	Y / 7	Y	Y / NA	N	NA	NA	22.9 x 12.7 x 17.8; 0.80 x 5.65 x 1.08	2.50; 0.03
mBio	10	Y	Y / N	20	Y	SA / 8 – 10	Y	Y / NA	N	NA	NA	38.0 x 19.0 x 15.0; 0.9 x 0.3 x 0.1	3.00; 0.02
Burnet	30	Y	Y / N	40	N	SA / 10	Y [2] / N	N	N	NA	NA	12.0 x 7.7 x 8.5; 9.0 x 2.0 x 0.5	0.37; 0.03
Zyomyx	100	Y	Y / N	6	N	N / 10	N	N	N	NA	NA	10.0 x 1.4 x 1.4	0.06

[1] Where applicable, the *instrument's* dimensions and mass are given first, followed by the *testing kit's* values.

[2] The results can be read manually or an additional instrument platform can be linked to automatically score the test.

Abbreviations: Y, yes; N, no; SA, semi-automated; NA, not available; FDA, Food and Drug Administration; CE, Conformité Européene; L, length; W, width; H, height.

quality control indicators for these three aspects.

When evaluating the characteristics of the PointCareNOW™, the CyFlow® miniPOC and the Alere PIMA™ as summarised in Table 2.2 against the desired characteristics for a POC CD4 device as summarised in Table 2.3, it becomes evident that none of these devices posess all of the desired characteristics. However, though the Alere PIMA™ does not comply with all of the requirements set out, the fact that it uses a capillary blood sample, is based on simple technology and has quality control procedures in place, makes it the technology that is most ready for implementation in low-resource settings. This is why the Alere PIMA™ CD4 testing technology was selected for use in this study.

Table 2.3: Interpretation of WHO ASSURED criteria for POC CD4 technology.

Wu & Zaman (2012) interpretation of ASSURED criteria for POC CD4 testing in low-resource settings	**Desirable characteristics for a POC CD4 test for low-income countries as defined by Zachariah *et al.* (2011)**
Affordable	
Less than US$500 per testing device.	
Less than US$10 per test.	Between US$1 and US$2 per test.
Sensitive and specific	
Lower limit of CD4+ detection 350 cells/mm^3.	
	The test should be robust and reliable so that repeat testing is not required.
User-friendly	
The device should be easy to use.	The test should be simple to use. (Dip stick or lateral flow technology is preferable.)
No more than one to two days of training should be required.	
	Use non-venous blood so that specialist skills (such as phlebotomy) are not required for sample collection.
	Results should be easy to read.
Rapid and robust	
Less than 90 minutes testing time.	Testing time should be short (preferrably 10 minutes).

Continued on next page

Continued from previous page

Wu & Zaman (2012) interpretation of ASSURED criteria for POC CD4 testing in low-resource settings	**Desirable characteristics for a POC CD4 test for low-income countries as defined by Zachariah *et al.* (2011)**
Reagent / kit shelf life of more than one year at room temperature.	The test kits and consumable should have a shelf life of at least fifteen months and should ideally be able to withstand temperatures of up to 40°C.
A minimal amount of consumables should be required.	
A high throughput is required.	
Equipment-free	
A compact, battery powered device.	If possible, no electronic instrumentation should be required.
Data analysis to be conducted on-site.	
Easy sample handling.	
No cold chain required for reagents or kits.	The test kits and consumable should not require a cold chain.
Easy disposal of all waste that is generated.	It should be easy to dispose of all waste safely.
Deliverable to end-users	
The device should be portable and preferrably hand-held.	
	It must be possible to adapt the test to a quality assurance programme that can ensure that tests are functioning properly and results are reliable.

The performance of the Alere PIMA™ CD4 test has been evaluated in studies by Mwau *et al.* (2013), Glencross *et al.* (2012), Manabe *et al.* (2012), Sukapirom *et al.* (2011) and Mtapuri-Zinyowera *et al.* (2010). All of these studies compared the performance of the Alere PIMA™ to that of gold standard laboratory-based CD4 testing. The findings differed across the various studies, with some studies classifying the Alere PIMA™'s performance as satisfactory and other studies recommending that the technology should be refined or further quality control mechanisms should be incorporated before the device will be ready for use in the field. The most important findings are summarised below:

1. Manabe *et al.* (2012) found that CD4 cell counts were negatively biased at higher absolute numbers and noted that this will limit the test's usefulness as a monitoring device for patients on ART. Sukapirom *et al.* (2011) also found the CD4

cell count obtained by the Alere PIMA™ to be lower than that obtained by the gold standard technology and agrees with Manabe *et al.* (2012) that this occurs more often in samples with a higher absolute CD4 count. Sukapirom *et al.* (2011) did not find this negative bias to be statistically significant. Mwau *et al.* (2013) also found the CD4 cell counts to be lower than those measured by the laboratory gold standard, but this finding was not limited to higher absolute CD4 cell counts. Glencross *et al.* (2012) found positive bias during rural clinic field studies, but noted that, with additional support from the manufacturer, this bias was reduced during a subsequent inner-city clinic field study;

2. Mwau *et al.* (2013) found the precision of the Alere PIMA™ to be unacceptable and the repeatability of the test to be less than the rate reported by other evaluations of the technology and recommended that the technology should be further refined to address these problems. Sukapirom *et al.* (2011) found the Alere PIMA™ to give "highly precise and reproducible results". Glencross *et al.* (2012) concluded that the variable precision of the Alere PIMA™ test during various stages of the study could be attributed to variable capillary sampling techniques and recommended that adherence to the specified capillary sampling technique at the POC is of paramount importance. The authors noted that the development of a capillary sampling quality control method by Alere was being awaited. Mwau *et al.* (2013), Manabe *et al.* (2012) and Mtapuri-Zinyowera *et al.* (2010) found that using the Alere PIMA™ with either venous blood samples or finger pricks (capillary sampling) yielded satsifactorily similar results;

3. Mtapuri-Zinyowera *et al.* (2010) found that there was no significant difference in the results obtained when the test was performed by a laboratory technician versus when it was performed by a nurse; and

4. Sukapirom *et al.* (2011) noted the fact that the test can only provide the absolute CD4+ T-lymphocyte count rather than the frequency of the CD4+ T-lymphocyte count as a limitation and noted that it would be good if the technology can be expanded to include this capability in the future.

As is evident from the discussion on the performance evaluation studies being conducted on the Alere PIMA™ above, the device is currently being evaluated in pro-

gramme settings. As such, data gathering on the cost of using the device is on-going. Larson *et al.* (2012) developed a standardised methodology for estimating the cost per test for POC CD4 testing from the provider's perspective so that data from different studies can be compared accurately. As mentioned in Section 1.2, studies in South Africa have found that POC CD4 testing is more expensive than traditional laboratory-based testing. The difference in the cost per test for laboratory-based CD4 testing and POC CD4 testing is quantified in Subsection 5.1.4.

2.3.4 Possible changes to the current service delivery model that are to be investigated

South Africa has more than 7 000 DoH public healthcare facilities, including hospitals, clinics and community healthcare centres, where services are provided to patients. ART initiation services are available at approximately 3 600 of these sites.

The country has more than 300 public healthcare laboratories (NHLS laboratories) that are responsible for providing diagnostic testing services to the patient-serving DoH facilities, 61 of these NHLS laboratories offer CD4 testing.

The purpose of this study is to investigate whether POC CD4 testing should be used in South Africa and how it should be incorporated into the diagnostic service provision model (if it is to be used). In order to answer this question, it is necessary to investigate the financial impact and the health impact of a number of different scenarios – these scenarios should include a scenario where no POC CD4 testing is used, a scenario where no laboratory-based CD4 testing is used as well as a number of scenarios where a combination of POC CD4 testing and laboratory-based testing is used. For each scenario, the following decisions must be made:

1. For each of the ART initiation facilities, a decision must be made to either provide POC CD4 testing capability at the healthcare facility itself, or not; and

2. For each ART initiation site where POC testing capability will not be provided, it is necessary to determine to which of the existing CD4 laboratories the facility should refer its CD4 samples.

2.4 Conclusion: The real-world problem

This chapter described the real-world problem by providing background on the current state of healthcare delivery in South Africa as well as the medium-term goals and objectives for public healthcare in the country. The current diagnostic service delivery model was discussed, with emphasis placed on the current role of POC testing in HIV and TB treatment. Finally, the Alere PIMA™ device for CD4 testing was introduced and the possible changes to the diagnostic service delivery model that are to be explored in this study were defined.

The following chapter will discuss the role of Operations Research in supporting decision-making within healthcare and will introduce the field of location science.

Chapter 3

Operations Research and location science in healthcare

The real-world problem was introduced in Chapter 2 through a description of the state of healthcare in South Africa, including the diagnostic service delivery model.

This chapter describes the role of Operations Research in supporting decision-making within the healthcare sector and introduces the field of location science and its application to healthcare.

3.1 Operations Research and healthcare

3.1.1 Operations Research

The science of Operations Research is a well-established field, academically as well as in industry. Here, the field is briefly introduced through a definition of Operations Research as well as a description of the most important concepts that are necessary for understanding the remainder of this document.

The field of Operations Research can be traced back to early attempts to incorporate scientific methods into the management of organisations. The name Operations Research was first used during the Second World War when mathematical modelling and optimisation techniques were employed in the management of military services. During the post-war industrial boom, the same types of questions that had been answered in a military context arose in industry, and the field of Operations Research grew as new applications were developed (Hillier & Lieberman, 2010).

Winston (2004) gives the following definition for Operations Research: "Operations Research (often referred to as management science) is simply a *scientific approach to decision-making* that seeks to best design and operate a *system*, usually under conditions requiring the allocation of scarce resources." Winston (2004) describes a *system* as an organisation of interdependent components that work together to establish a common goal. The *scientific approach to decision-making* usually entails the use of one or more mathematical models (Winston, 2004).

The mathematical models used have three key components: (i) the objective function(s); (ii) the decision variable(s); and (iii) the constraint(s). Optimisation is the process of searching for the value(s) of each of the decision variables that simultaneously satisfy all of the given constraints and optimise (maximise or minimise) the value(s) of the objective function(s) (Winston, 2004).

The mathematical models used in Operations Research can be classified according to various criteria. Some of the most commonly used classifications that are relevant to subsequent discussions in this document are (Winston, 2004):

1. Static vs. Dynamic: In static models, the values of coefficients and constants in the objective function and constraint equations are assumed to remain fixed over time. In dynamic models, these values do not remain constant over time;

2. Deterministic vs. Stochastic: In deterministic models, the values of coefficients and constants in the objective function and constraint equations are assumed to be known with certainty. In stochastic models, these values are represented by statistical or empirical distributions because they are not known with certainty;

3. Continuous vs. Discrete: In continuous models, decision variables may assume any (non-integer) value, in discrete models, decision variables may only assume set (usually integer) values; and

4. Single-objective vs. Multi-objective: Multi-objective optimisation models contain two or more conflicting objectives. For this reason, no single, optimal solution to a multi-objective optimisation model exists. Rather, the optimal solution set is a collection of solutions known as a Pareto set. A solution is a member of the Pareto set if it is not possible to improve the performance of the solution in terms

of any one of the objectives without decreasing its performance in terms of at least one of the other objectives.

The process of optimising (solving) a mathematical model may require a large amount of computational effort. The development of different solution methods for mathematical models represents an independent research field within Operations Research. At present, three main solution methods exist:

1. Exact Algorithms: An exact algorithm is a defined search procedure that guarantees that the solution that is generated is the optimum answer to the problem. The use of an exact algorithm can become infeasible due to the computational effort required for certain mathematically complex problems. Exact algorithms can not be applied to problems with more than one objective function;

2. Heuristics: Heuristics are most often developed with the purpose of solving a specific problem efficiently. Like exact algorithms, heuristics also define a search procedure, however, unlike exact algorithms, there is no guarantee that the answer that is generated is in fact the optimum answer to the problem (Hillier & Lieberman, 2010); and

3. Metaheuristics: A metaheuristic is a more high-level search procedure that may incorporate the use of heuristics. A metaheuristic incorporates feedback procedures that enables it to search through the potential solutions in an attempt to generate solutions that are closer to optimal as it goes along. In contrast to heuristics that are generally developed with the purpose of solving a specific problem type, metaheurstics can be used to solve a wider variety of problems. As with heuristics, there is no guarantee that the final solution that is proposed is the optimum solution to the problem. Metaheuristics are useful in solving problems that are too computationally burdensome to be solved by exact algorithms as well as problems with more than one objective function.

Computational complexity theory is concerned with quantifying the complexity of solving any given problem. The concept of polynomial time is one of the methods that are used to quantify computational complexity. The term "NP-hard" will be used later in this chapter, the following brief descriptions of complexity classifications are given in order to provide a sufficient understanding of the term "NP-hard":

1. A problem can be classified as **P** (i.e. as being of polynomial time) if it can be *solved* (by a theoretical computer, called a Turing machine) in a time that does not exceed a polynomial expression that has been defined in terms of the problem size;

2. A problem is classified as **NP** if it can be *verified* on a non-deterministic Turing machine in polynomial time. From this definition it follows that all problems that are classified as P can also be classified as NP. However it is unclear whether the inverse is true. This question (whether NP = P) is considered to be one of the most important questions that currently exist in mathematics (Anthonissen, 2007);

3. Cook (1971) proved the existence of subset of NP problems called **NP-complete** problems. A problem (call the problem Ψ) can be classified as NP-complete if any NP problem can be reduced to Ψ by a Turing machine in polynomial time; and

4. A problem is classified as **NP-hard** if it is at least as computationally complex as the most complex NP problems. More formally, a problem (call the problem Ω) can be classified as NP-hard if there exists an NP-complete problem that can be reduced to Ω by a Turing machine in polynomial time (Anthonissen, 2007). NP-complete problems are a subset of NP-hard problems. NP-complete problems include all problems that can be classified as both NP and NP-hard. The set of NP-hard problems also contains problems that can not be classified as NP.

The relationship between the various classifications described here is illustrated in Figure 3.1. The diagram on the left represents the relationship if it were to be proven that $P \neq NP$, while the diagram on the right illustrates the case if it were to be proven that $P = NP$.

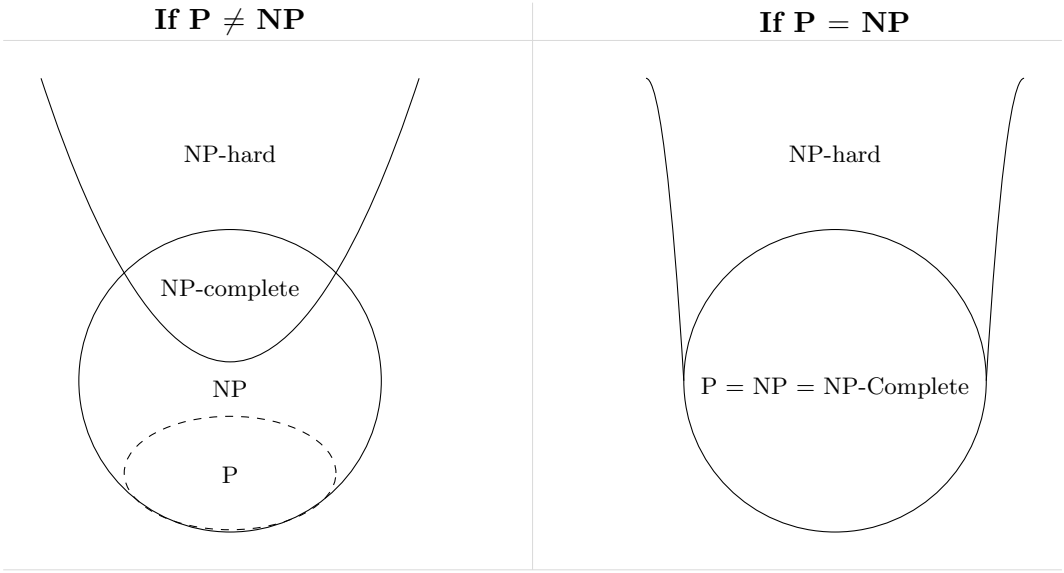

Figure 3.1: Relationship between P, NP, NP-complete and NP-hard. (Diagram conceptualised by Benham Esfahbod.)

3.2 The need for Operations Research in healthcare

Simon Fraser University's Complex Systems Modelling Group (2010) states that, worldwide, there is an increasingly popular view that conventional approaches to the design of healthcare systems are failing. This growing realisation is attributed to the increased pressure that healthcare providers are facing as they attempt to continue providing quality services while maintaining control over escalating costs. The book describes the questions facing the healthcare sector as extremely complex and states that, in the majority of cases, it is not sufficient to rely on intuition to answer a given question. The Complex Systems Modelling Group (CSMG) describes an increased use of modelling as a method of either (i) developing evidence-based answers to the complex questions facing healthcare decision-makers; or (ii) characterising a problem and exploring different ways to resolve it in order to help steer healthcare decision-makers in the right direction.

Reid *et al.* (2005) confirm the CSMG's assertion when it describes "a growing realization within the healthcare community of the critical role information and communications technologies, systems engineering tools, and related organisational innovations must play in addressing the interrelated quality and productivity crises facing the

healthcare system". This book describes the contrast between the instrumental role that engineering plays in the development of medical technologies and the support of medical research and the very limited application of engineering techniques to the design and management of the healthcare *delivery* system. It describes the complex nature of healthcare delivery systems (involving interrelated systems of distributed, specialised personnel, multiple information and material flows, specialised care facilities, financial resources, etc.) as well as the similarities between these complexities and those found in other systems where systems thinking and industrial engineering techniques have successfully been applied to improve operations. Reid *et al.* (2005) summarises the problem as follows: "In short, the principles, tools, and research from engineering disciplines associated with the analysis, design, and control of complex systems (Systems Engineering, Industrial Engineering, Operations Research, Human-factors Engineering, Financial Engineering/Risk Analysis, Materials/Microelectromechanical Systems Engineering, etc.) – disciplines that have helped improve, and sometimes transform, many manufacturing and other services industries – are largely unknown in the clinical operations of healthcare delivery."

Brandeau *et al.* (2004) group the decisions facing healthcare policy makers and planners into two broad categories: (i) healthcare planning and organising; and (ii) healthcare delivery. Healthcare planning and organising is defined as encompassing high-level decision-making and includes decisions on the economics and structure of healthcare systems while healthcare delivery encompasses lower-level decision-making that is mainly concerned with the management of healthcare operations and with clinical practice. This structure is depicted in Figure 3.2.

This study fits within the "Operations management for healthcare delivery" domain. The operations management problems that arise within healthcare are similar to traditional operations management problems and include the following:

1. Strategic planning problems (e.g. design of services);
2. Design of the healthcare supply chain;
3. Facility planning and design;
4. Equipment evaluation and selection; etc.

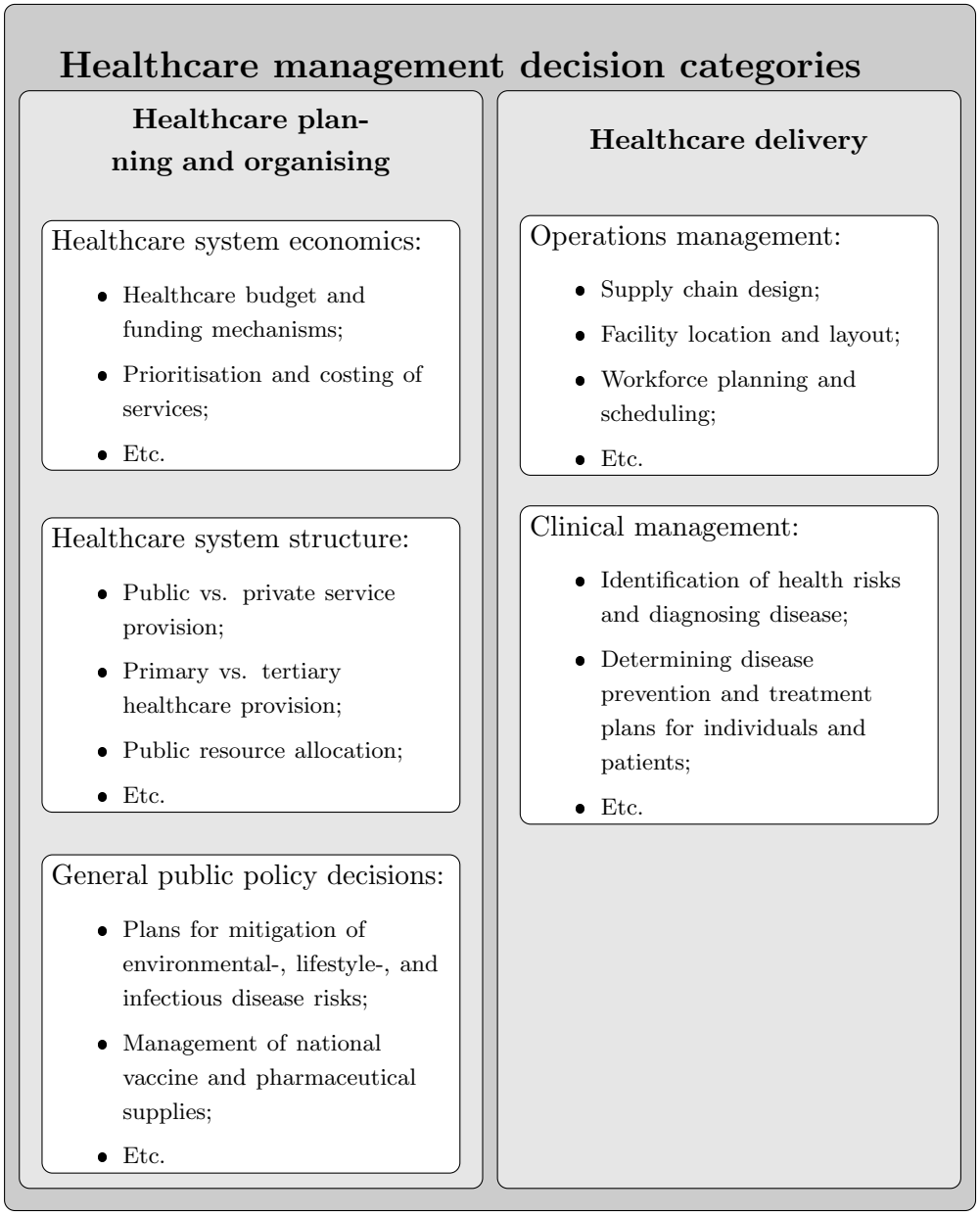

Figure 3.2: Categories of healthcare management decision-making.

According to the description of the real-world problem in the preceding chapters, there is a need to develop scenarios for incorporating POC testing into the existing diagnostic service provision model by determining (i) which hospitals and clinics to place POC CD4 devices at; and (ii) which CD4 testing laboratories the remaining hospital and clinic sites should refer their work to. From this description, it is evident that the field of location science (which fits within the Operations Research domain) should be explored as a possible approach to solving the real-world problem.

3.3 Location science

In a review of location science research, Hale & Moberg (2003) give the following description for facility location problems: "Facility location problems investigate where to physically locate a set of facilities (resources) so as to minimise the cost of satisfying some set of demands (customers) subject to some set of constraints." ReVelle & Eiselt (2005) describe four components that characterise a location problem:

1. *Customers* (it is assumed that these customers are already located at specific points or on specific routes);
2. *Facilities* that need to be located;
3. *A space* wherein customers are located and facilities are to be located; and
4. *A metric* that indicates distances or travel times between customers and facilities.

When reviewing the list of decisions (described in Subsection 2.3.4) that need to be made to solve the real-world problem under investigation, against the definition given by Hale & Moberg (2003) and the key components defined by ReVelle & Eiselt (2005), it becomes apparent that location science could be a suitable approach for solving the real-world problem. When fitting the real-world problem to the given definition for location science:

1. The *facilities to be located* would be the allocation of POC CD4 testing capability to an ART initiation site or the allocation of an ART initiaiton site to an existing CD4 laboratory; and

2. The *customers* or *set of demands that need to be satisfied* would be the number of CD4 test requests generated at each patient-serving healthcare facility (hospital, clinic, etc.).

When assessing the real-world problem against the four key components of a location problem:

1. The *customers* (the patient-serving healthcare facilities) are located at specific points;
2. There are a set of *facilities* (CD4 testing capability) that need to be located;
3. A defined *space* exists wherein customers are located and facilities need to be located (the existing public healthcare hospitals, clinics, community healthcare centres and CD4 laboratories in the country); and
4. Either the *distance* or the *travel time* between each of the *customer* and *facility* nodes in the country can be calculated.

Location science is a well-established research area within the field of Operations Research and several researchers have published textbooks and articles that provide detailed accounts of the field and its applications. It is impossible to summarise the literature here, rather, the aim is to provide a general overview of the field of location science, its development, research gaps that currently exist, industries that location science have been applied to and the use of location science in healthcare-related problems.

3.3.1 A brief overview of the field of location science

Hale & Moberg (2003) state that location science as an area of analytical study can be traced back to the early seventeenth century when Pierre de Fermat, Battista Cavalierri and Evagelistica Torricelli all independently proposed the basic Euclidian spatial median problem (also known as the Weber problem). ReVelle & Eiselt (2005) credit Torricelli with the first literature on a location problem. Owen & Daskin (1998) argue that the formal study of location theory began in 1909 when Alfred Weber studied the positioning of a single warehouse in such a way that the total distance between the warehouse and its customers would be minimised (this is the Weber problem referred

to by both Hale & Moberg (2003) and ReVelle & Eiselt (2005)). In the period between the early 1900s and the 1960s, the research focus remained on locating a single facility within a two-dimensional field (ReVelle & Eiselt, 2005). From the 1960s onward, the research focus widened to include problems that considered the location of several facilities and this lead to increased research interest in the field (Owen & Daskin, 1998). Hale & Moberg (2003) describe near linear growth in the annual number of location science study publications between the 1960s and the 2000s.

Current *et al.* (2002) attribute the sustained interest in location science to the following five factors:

1. Location decisions are made at all levels of human organisation – ranging from individuals and households, through commercial organisations and up to international governing agencies;

2. Location decisions are most often strategic in nature as they typically require significant financial resources and have a long lifetime. In the public sector, location decisions significantly impact the ability to deliver services to the population whilst in the private sector, location decisions typically impact a company's ability to compete in a given market place;

3. Location decisions often have a wide-scale impact on economic, social, environmental and health factors, amongst others;

4. Location problems are typically difficult to solve (with a large number of classic model types being classified as NP-hard); and

5. Location problems tend to be application-specific with the objectives, the constraints and the variables (the structural form) being determined by the specific problem being investigated.

Earlier, reference was made to the large volume of location science literature that has been produced. However, both ReVelle & Eiselt (2005) and Current *et al.* (2002) state that, though there is a wealth of literature available on different models (mathematical formulations) for location science problems as well as on efficient solution techniques for these models, there is a significant shortage of literature that describes the practical application of location science to real-world problems. When one recalls

the application-specific nature of location science problems described by Current *et al.* (2002), it becomes apparent that this lack of literature on location science applications to real-world problems represents a meaningful research gap.

ReVelle & Eiselt (2005) attribute the shortage of application literature to the difficulties associated with fitting complex, strategic, real-world problems that often involve multiple objectives to the theoretical models (mathematical formulations) that have been developed. The authors suggest either developing more comprehensive models that can deal with these complexities holistically or developing models that consist of a number of subproblems and using extensive sensitivity analysis to analyse and synthesise the subsets of results that are obtained. ReVelle & Eiselt (2005) also suggest that integrating Geographic Information Systems (GIS) with location science so that multiple decision-makers can work with an input interface and see the results from input modifications to the model graphically displayed, may increase the use of location science in real-world applications.

Location science is studied in a large variety of academic disciplines including Operations Research, Industrial Engineering, Civil Engineering, Geography, Urban Planning and Supply Chain Management (Hale & Moberg (2003), Dimopoulou & Giannikos (2007)).

3.3.2 Location science model types

Literature contains different opinions on what the basic location science model types are, however, the following eight basic model types for networks are most commonly defined:

1. The set covering problem – ReVelle & Eiselt (2005), Hale & Moberg (2003) and Current *et al.* (2002);

2. The maximum (or *maximal*) covering location problem – ReVelle & Eiselt (2005), Hale & Moberg (2003) and Current *et al.* (2002);

3. The P-center problem – ReVelle & Eiselt (2005), Hale & Moberg (2003) and Current *et al.* (2002);

4. The P-median problem – ReVelle & Eiselt (2005), Hale & Moberg (2003) and Current *et al.* (2002);

5. The uncapacitated facility location problem (including *fixed charge location problems*) – ReVelle & Eiselt (2005) and Current *et al.* (2002);
6. The maxisum location problem Current *et al.* (2002) (referred to as the "undesirable facility location problem" and defined as an extension of the basic problem structures by ReVelle & Eiselt (2005));
7. The P-dispersion problem – Current *et al.* (2002); and
8. The hub location problem – Current *et al.* (2002) (defined as an extension to the basic problem structures by ReVelle & Eiselt (2005)).

(Note: the "P" used in the names of the basic location science model types should not be confused with the "P" used in the classification of complexity.)

Daskin & Dean (2004) state that the majority of facility location models are derived from only three of the basic models listed above: (i) the location set covering model; (ii) the maximal covering model; and (iii) the P-median model. In addition to the two examples included in the list above, ReVelle & Eiselt (2005) define the following seven extensions to the basic problem structures:

1. Cent-Dians and Medi-Centers;
2. Balancing objectives;
3. Hierarchical siting problems;
4. Hub location problems;
5. Competitive location problems;
6. Combined siting and routing; and
7. Capture problems.

The eight basic model types will be discussed in more detail in Chapter 4 and the standard mathematical notation for each model will be given. The seven extensions to the basic model structures will not be discussed in any further detail, the interested reader is referred to ReVelle & Eiselt (2005) for a detailed discussion of these.

3.3.3 Applications of location science

Location science is applied to a wide range of problem types. Various authors have reviewed the literature on real-world applications of location science: Current *et al.* (2002) compiled an overview of general applications; Ho (2008) reviewed applications where traditional Operations Research modelling techniques were integrated with the "analytical hierarchy approach" method; Melo *et al.* (2009) provided an overview of applications to supply chain optimisation; Farahani *et al.* (2012) reviewed covering problems; and Boloori Arabani & Farahani (2012) considered only dynamic models. Some of the examples given by each of these authors are summarised in Table 3.1.

Table 3.1: Examples of applications of location science.

Application	**Citation**[1]
Advertising medium selection	Kwak et al. (2005)[1]
Agriculture harvesting measure selection	Guo and He (1999)[1]
Air craft alert site location	Bell et al. (2011)[3]
Airline hubs	O'Kelly (1987)[2]
Airports	Saatcioglu (1982)[2]
Apparel sizing	Tryfos (1986)[2]
Archaeological settlement analysis	Bell and Church (1985)[2]
Auto emission testing stations	Swersey and Thakur (1995)[2]
Automotive supply chain	Fleischmann et al. (2006)[4]
Bank account location	Cornuejols et al. (1977)[2]
Bank branch location	Alexandris and Giannikos (2010)[3]
Bloodbank	Price and Turcotte (1986)[2]
Brewery depots	Gelders et al. (1987)[2]
Bus garages	Maze et al. (1981)[2]
Bus stops	Gleason (1975)[2]
Chemicals supply chain	Jayaraman and Ross (2003)[4]
Chip manufacturing	Cho and Sarrafzadeh (1994)[2]
Coal handling facilities	Osleeb and Ratick (1983)[2]
Computer service centers	Ghosh and Craig (1986)[2]
Customer data collection method selection	Badri (2001)[1]
Damper placement	Kincaid and Berger (1993)[2]
Database management	Pirkul (1986)[2]
Daycare centers	Holmes et al. (1972)[2]
Electric power generating plants	Cohon et al. (1980)[2]
Electronics industries	Gebennini et al. (2009)[5]
Emergency equipment for oil spills	Belardo et al. (1984)[2]
Emergency medical services	ReVelle et al. (1977)[2]
Environment land use pattern selection	Malczewski et al. (1997)[1]

Continued on next page

Continued from previous page

Application	Citation[1]
Essential air services	Flynn and Ratick (1988)[2]
Evacuation management (dynamic model)	Mamada et al. (2005)[5]
Fast-food restaurants	Min (1987)[2]
Fire station systems (dynamic model)	Yang et al. (2007)[5]
Fire stations	Schilling et al. (1980)[2]
Flexible manufacturing system tool selection	Daskin et al. (1990)[2]
Food supply chain	Avittathur et al. (2005)[4]
Forest harvesting sites	Hodgson et al. (1987)[2]
Forestry supply chain	Vila et al. (2006)[4]
Franchise outlets	Pirkul et al. (1987)[2]
Government internet access technology selection	Malladi and Min (2005)[1]
Grain subterminals	Hilger et al. (1977)[2]
Hardware supply chain	Laval et al. (2005)[4]
Hazardous waste disposal sites	ReVelle et al. (1991)[2]
Healthcare IT-based project selection	Kwak and Lee (2002)[1]
Higher education faculty course assignment	Ozdemir and Gasimov (2004)[1]
Ingot size selection	Vasko et al. (1988)[2]
Landfill industries (dynamic model)	Melachrinoudis et al. (1995)[5]
Logistics subcontractor selection	Lee and Hsu (2004)[1]
Logistics supplier selection	Ghodsypour and O'Brien (1998)[1]
Manufacturing/distribution system (dynamic model)	Melachrinoudis and Min (2000)[5]
Manufacturing maintenance strategy selection	Bertolini and Bevilacqua (2006)[1]
Manufacturing material handling device selection	Braglia et al. (2001)[1]
Manufacturing sub-component selection	Akgunduz et al. (2002)[1]
Meat industries	Schütz et al. (2008) [5]
Medical diagnosis	Reggia et al. (1983)[2]
Metallurgical grade assignment	Vasko et al. (1989)[2]
Military nuclear fuel cycle selection	Kim et al. (1999)[1]
Military ship selection	Crary et al. (2002)[1]
Military supply chain	Farahani and Asgari (2007)[4]
Oil industries (dynamic model)	Mokhtarian (2011)[5]
Paper recycling supply chain	Pati et al. (2008)[4]
Police patrol area design	Curtin et al. (2010)[3]
Political party platform	Ginsberg et al. (1987)[2]
Product positioning in feature space	Gavish et al. (1983)[2]
Product procurement and standardisation	Watson (1996)[2]
Production lot sizing	Van Oudheusden and Singh (1988)[2]
Public library policies (dynamic model)	Koontz et al. (2009)[5]
Public swimming pools	Goodchild and Booth (1980)[2]
Railroad sidings	Higgins et al. (1997)[2]
Rain gauges	Hogan (1990)[2]
RFID network design for asset tracking in healthcare	Oztekin et al. (2010)[3]

Continued on next page

Continued from previous page

Application	Citation[1]
Rural healthworkers	Bennett et al. (1982)[2]
Sand supply chain	Liste and Dekker (2005)[4]
Satellite homing stations	Helme and Magnanti (1989)[2]
Satellite orbits	Drezner (1988)[2]
Schools (dynamic model)	Antunes and Peeters (2001)[5]
Social service centers	Patel (1979)[2]
Solar power system design	Birge and Malyshko (1985)[2]
Solid waste collection	Marks and Liebman (1971)[2]
Telecommunication switching centers	Hakimi (1965)[2]
Truck terminals	Love et al. (1985)[2]
Transportation (dynamic model)	Bozkaya et al. (2010)[5]
Vehicle inspection stations	Hodgson et al. (1996)[2]
Vehicle routing	Bramel and Simchi-Levi (1995)[2]
Warehouses	Kuehn and Hamburger (1963)[2]
Wi-Fi equipment location	Lee and Murray (2010)[3]

[1] Citations are as given by Ho (2008)
[2] Citations are as given by Current *et al.* (2002)
[3] Citations are as given by Farahani *et al.* (2012)
[4] Citations are as given by Melo *et al.* (2009)
[5] Citations are as given by Boloori Arabani & Farahani (2012)
These citations have not been included in the references of this document as they were not directly referenced in the writing of this thesis.

In the next section, this discussion is continued with a more detailed focus on the application of location science specifically to healthcare-related problems.

3.3.4 Location science and healthcare

Daskin & Dean (2004) classify healthcare and related location literature into three broad categories:

1. Accessibility models: Accessibility models use static inputs (e.g. the demand at a node, the travel time between nodes, and the cost of establishing a facility are often assumed to be fixed and to remain constant over a time period) to determine the optimal placement of healthcare facilities. Due to the static inputs that are used, accessibility models are frequently applications or limited extensions of the basic location science model types defined in Subsection 3.3.2. Accessibility models also include hierarchical models where different levels of service provision (e.g. primary healthcare clinics, district hospitals and tertiary hospitals) are to be located in a tiered network. Lastly, though Daskin & Dean (2004) state that

the vast majority of healthcare location science models are discrete, there are also some examples of continuous accessibility models in literature;

2. Adaptability models: Adaptability models attempt to locate facilities under conditions of uncertainty regarding the future (e.g. the demand that will be present at each node at some point in the future may be unknown). These models often make use of scenario analysis in order to make a recommendation on the best way forward. Under conditions of uncertainty, the recommended ("best") solution is not necessarily the optimal solution for a particular scenario, rather, it is the solution that fares best across all of the scenarios. (The term "regret" refers to the difference between the performance of the optimal solution and the performance of the recommended solution for any one scenario.) In scenario planning, certain decisions (e.g. the placement of healthcare facilities) need to be made before it is known which scenario will eventually come into being, whilst other decisions (e.g. the assignment of medical staff to the healthcare facilities) can be made after the true scenario is revealed. Scenario planning typically makes use of the following three performance measures: optimising the expected performance, minimising the worst case performance and minimising the worst case regret; and

3. Availability models. Availability models are concerned with short-term uncertainty regarding the operating conditions of a system. (For example, an ambulance may have no emergencies occuring in the area which it is assigned to for an extended period of time and then have two emergencies occuring which it needs to respond to simultaneously.) There are two approaches to dealing with this short-term uncertainty:

   (a) Deterministic: Models are set up in such a way that demand nodes are covered multiple times (whenever possible). For example, in cases where multiple optimal solutions exist, the solution that maximises the system-wide multiple coverage is selected, or, in cases where multiple optimal solutions do not exist, a constraint may be added that requires that each demand node be covered at least twice, etc.;

   (b) Stochastic (probabilistic): Models make use of queueing theory, Bernoulli trials, etc. to account for the uncertainty that is present in the model.

Table 3.2 summarises examples of the application of location science to healthcare problems.

Table 3.2: Examples of applications of location science to healthcare.

Application	Description	Citation
Accesibility models		
Healthcare facilities (Chongju Metropolitan Area, Korea)	A multi-objective equity-efficiency trade-off model that incorporates various basic model types, including the two-parameter production-constrained gravity model, is developed. The Monte Carlo integer programming technique is combined with the augmented Lagrangian algorithm to solve the case study.	Cho (1998)
Healthcare facilities, hierarchical placement (Rio de Janeiro, Brazil)	A three-level successively inclusive hierarchichal model is applied to the case study (which is focused on maternal and perinatal facilities). Several relaxations and heuristics for solving the problem are developed and tested.	Galvao *et al.* (2002)
Emergency medical services locations (Theoretical development of modeling and solution methodology)	A maximal covering location model is expanded to include three objectives. Two solution approaches are tested: lexicographic linear programming; and fuzzy goal programming.	Araz *et al.* (2007)
Healthcare facilities, hierarchical placement (South India)	Three maximum cover type models and three P-median type models are implemented on case study data and compared. A new mathematical model, specifically designed to locate the maximum number of sustainable facilities is developed and applied to a case study.	Smith *et al.* (2009)
Long-term care facilities locations (Chuncheong Provinces, Korea)	A customised mathematical model is developed for use in siting long-term care facilities (typically for the elderly). A customised solution heuristics is developed and tested on a case study.	Kim & Kim (2010)
Healthcare facilities (Malaysia)	The capacitated maximal covering location problem is applied to a case study. A new solution algorithm (based on the genetic algorithm) is proposed and applied.	Shariff *et al.* (2012)
Healthcare facilities (urban sub-districts, China)	An extension of the cost-based P-median location-allocation model is applied to the case study (which is focussed on non-communicable disease prevention and control units). The efficiency of non-inferior solutions is evaluated based on their relative cost and service.	Guo *et al.* (2013)

Continued on next page

Continued from previous page

Application	Description	Citation
Healthcare facilities (Guadalajara and Cuenca Provinces, Spain)	A customised multi-objective mathematical formulation (with distance as one objective and cost as the other) is developed for the case study (which looks at facilities that provide 24-hour care) and solved using a harmony search algorithm.	Landa-Torres *et al.* (2013)
Adaptibility models		
Road network design for improved access to healthcare facilities (Suhum District, Ghana)	A new mathematical model (the maximal covering network improvement problem) is formulated and applied to a case study. Several different scenarios are modelled and the results are compared.	Murawski & Church (2009)
Healthcare facility locations and road network design for improved access (Nouna district, Burkina Faso)	The budget facility location-network design problem is modeled as a mixed integer programme and applied to a real-world problem. Several different scenarios are modelled and the results are compared.	Cocking *et al.* (2012)
Emergency blood supply scheduling (Beijing, China)	A multi-period location-allocation model is applied to the case study and a heuristic algorithm (based on Lagrangian relaxation) is developed and applied to the case study.	Sha & Huang (2012)
Healthcare facilities (Illam Province, Iran)	The budget-constrained dynamic uncapacitated facility location-network design problem is developed and applied to a case study. Two heuristics are proposed for solving the problem.	Ghaderi & Jabalameli (2013)
Reduction and consolidation of healthcare facilities (Peloponnese Province, Greece)	An iterative three-step process incorporating data envelopment analysis and integer programming location-allocation models is applied to evaluate scenarios in a case study.	Mitropoulos *et al.* (2013)
Availability models		
Real-time emergency medical services reallocation (Montreal, Canada)	A dynamic reallocation model and a parallel Tabu search heuristic to calculate reallocation scenarios before they are required is developed and applied to a case study.	Gendreau *et al.* (2001)
Emergency medical services locations (Theoretical model was developed and applied to a classical reference problem)	Stochastic elements are incorporated separately into each stage of a two-stage model in the form of probabilities in the contraint equations.	Beraldi *et al.* (2004)

Continued on next page

Continued from previous page

Application	Description	Citation
Emergency medical services locations (Riyadh, Saudi Arabia)	The maximal covering location problem is expanded to incorporate stochastic elements and applied to a case study. Goal programming is used to accommodate more than one objective.	Alsalloum & Rand (2006)
Emergency medical services locations (Theoretical development of modeling and solution methodology)	The capacitated facility location problem is expanded to incorporate stochastic elements (when emergencies will occur and the travel time to emergencies). A two-stage stochastic model is developed that incorporates probabilities in the constraint equations. An exact solution method and three heuristics for solving the model are developed and tested.	Beraldi & Bruni (2009)
Emergency medical services locations (Ground and air ambulances, New Mexico, USA)	Extensions of both the location set covering problem and the maximal covering location problem are applied to a case study and solved using a greedy heuristic.	Erdemir *et al.* (2010)
Real-time emergency medical services reallocation (Athens, Greece)	A new two-step model is developed that incorporates the hypercube model and a location model. The new model is applied to a case study and solved.	Geroliminis *et al.* (2011)
Emergency medical services locations (Theoretical development of modeling and solution methodology)	Uncertainty regarding where emergencies will occur is addressed by setting the objective to minimise the expected maximum distance to emergencies. A solution method that is based on simulation is developed. The problem is continuous, on the plane.	Canbolat & von Massow (2011)
Emergency medical services locations (Rural outskirts of Tehran, Iran)	The linear location set covering problem is extended to form the linear reliability-based model. This model is applied to a case study in a rural setting.	Shariat-Mohaymany *et al.* (2012)
Emergency medical services locations (South East Wales, UK)	A new mathematical formulation (the maximal expected survival location model for heterogeneous patients) is developed, applied to a case study and solved using a genetic algorithm.	Knight *et al.* (2012)
Real-time emergency medical services reallocation (Vienna, Austria)	A customised mathematical model is developed specifically for the problem and applied to the case study. The problem is solved using approximate dynamic programming.	Schmid (2012)
Real-time emergency medical services reallocation (Waterloo, Canada)	A two-step stochastic model that aims to minimise the number of reallocations whilst maintaining an acceptable service level is developed and applied to a case study.	Naoum-Sawaya & Elhedhli (2013)

3.4 Conclusion: Operations Research and location science in healthcare

This chapter provided a brief introduction to Operations Research in general and to the field of location science in particular. The need for Operations Research in healthcare was identified and examples of both general and healthcare-specific applications of location science were given.

The following chapter will present the standardised mathematical formulations of location science problems before identifying the most suited formulation for application to the real-world problem.

Chapter 4

Mathematical models for location science applications

The previous chapter placed this study in context by providing background on (i) the role of Operations Research in healthcare; and (ii) location science and its application to healthcare.

In this chapter, the mathematical models that are typically used in location science, both in general applications as well as in healthcare-specific applications, are described in more detail and analysed. The discussion then moves on to the selection and adaptation of the base model for application to the real-world problem.

4.1 Mathematical formulations for general location science applications

Eight basic location science model types were introduced in Chapter 3. Literature contains several different criteria for categorising location science models. This section starts with a discussion of some of the proposed methods for classifying mathematical models before proceeding to a detailed discussion of each of the eight basic model types.

4.1.1 Classification of location models

The purpose of starting with a summary of various methods for classifying models is to make the reader more attuned to the subtle elements that differentiate models before presenting the eight basic models.

4.1.1.1 Classification according to objective

Several authors have suggested classifying the different types of location models according to their objective. Owen & Daskin (1998), Church (1999) and Hale & Moberg (2003) suggest defining four categories based on the overarching goal that the mathematical model seeks to achieve. Generally, the authors differed on the definition of the various categories. The three approaches that were proposed are summarised here.

Owen & Daskin (1998) proposed the following four categories:

1. Median problems – place a fixed number of facilities in a manner that minimises the average distance from any customer to its closest facility;
2. Covering problems – place facilities in such a way that all (the minimum required number of facilities to place is calculated as part of the model) or most (only a fixed number of facilities can be placed) of the demand in the system is covered within a specified distance;
3. Center problems – place a fixed number of facilities in a manner that minimises the maximum distance between the customer that is located furthest away from the facility to which it has been assigned; and
4. Other problem formulations.

Church (1999) agreed with Owen & Daskin (1998) on the definition of the first and second category only:

1. Median models – see description given in Owen & Daskin (1998) list;
2. Covering models – see description given in Owen & Daskin (1998) list;
3. Capacitated facility models – there is a limit on the number of customers that each faclity can serve; and
4. Competition models – location decisions can be adjusted (over predefined time frames) in response to the location decisions made by competitors.

The categories proposed by Hale & Moberg (2003) overlap with those that have been defined thus far:

1. Minisum – equivalent to the "median" category defined by both Owen & Daskin (1998) and Church (1999);

2. Minimax – equivalent to the "center problems" category defined by Owen & Daskin (1998);

3. Set covering – a subset of the "covering" category proposed by both Owen & Daskin (1998) and Church (1999). All demand in the model must be covered within a specified distance, thus the number of facilities to locate are determined as part of the solution; and

4. Maximal covering – also a subset of the "covering" category proposed by both Owen & Daskin (1998) and Church (1999). All demand in the model does not necessarily need to be covered. The number of facilities to locate are specified and the goal is to cover as much demand as possible within a specified distance.

Both Current *et al.* (2002) and ReVelle & Eiselt (2005) proposed more general approaches whereby models are classified based on their consideration of distance or their type of objective.

Current *et al.* (2002) defined the following three ways in which distance is considered in a location model's mathematical formulation:

1. Maximum distance (described as an "equity" objective) – encompasses the "covering problems" and the "center problems" defined by Owen & Daskin (1998);

2. P-dispersion problem – facilities are placed in such a way that the minimum distance between any pair of newly sited facilities is maximised; and

3. Total or average distance (described as an "efficiency" objective) – includes the "median" category defined by both Owen & Daskin (1998) and Church (1999) and the "capacitated facility models" category defined by Church (1999).

ReVelle & Eiselt (2005) essentially define the same three categories, but use different terminology:

1. Pull objectives – equivalent to the "maximum distance" category defined by Current *et al.* (2002);

2. Push objectives – equivalent to the "P-dispersion" category defined by Current *et al.* (2002); and

3. Balancing objectives (achieving equity) – equivalent to the "total or average distance" category defined by Current *et al.* (2002).

4.1.1.2 Classification according to other criteria

In addition to suggesting classifications based on the model objective (Subsection 4.1.1.1) both Hale & Moberg (2003) and ReVelle & Eiselt (2005) proposed alternative classification criteria. These proposals are summarised here.

Both groups of authors suggested using the space of location decisions to classify models into the following three categories:

1. Continuous – new facilities can be located anywhere within the space that is being modelled;
2. Discrete – new facilities can only be located at a number of predefined candidate sites; and
3. Network – new facilities can only be located on the arcs or nodes of a defined network.

ReVelle & Eiselt (2005) suggested a system for classifying models according to the number of facilities that are to be located. The following categories were defined:

1. One facility to locate; or
2. P facilities to locate, where:
   (a) P is given; or
   (b) P is not given ("free entry"). In instances where P is not given, P is either:
      i. Restricted by a budget; or
      ii. Determined endogenously through the objectives of the model.

Finally, ReVelle & Eiselt (2005) made two more suggestions: (i) a distinction could be made between models that *allow customers a choice* regarding the facility that they use (user attracting) versus those that *allocate customers to facilities* (delivery systems); and (ii) it is appropriate to distinguish between location science models that are developed for the public sector and those that are developed for the private sector.

4.1.2 Basic model types

The eight basic facility location models are described in detail in the following section. For each model, the key assumptions, the required model inputs and the model outputs are defined before the mathematical formulation is given.

4.1.2.1 Set covering problem

The set covering problem makes use of an indicator variable that indicates whether the demand at a specific node can be covered by a facility at a specific location based on the distance between the demand node and the facility. Two commonly used variations of the problem exist:

1. The *standard* set covering model that seeks to minimise the *total cost of locating* the selected facilities; and
2. The *location* set covering model that aims to minimise the *total number of facilities* that are located.

The *location* set covering model is typically applied when the operating costs (that are determined by the number of facilities that are located) are more dominant than the fixed cost of locating a facility (that is assumed to be approximately equal for all facilities (Daskin & Dean, 2004)).

The assumptions in this model are:

1. Facilities can only be located on the nodes of the network;
2. The demand is located at the nodes of the network;
3. The facility capacities are unlimited;
4. The demand nodes are unweighted;
5. The coverage distance can either be standardised across the model, or, additional restrictions on the set of facilities that can cover the demand at any particular node can be applied. (Additional restrictions would typically include factors such as the ease of travel from a demand node to a particular candidate location.);
6. All of the demand must be covered; and
7. The objective of both the *standard* set covering model and the *location* set covering model were described in the introductory paragraph above.

The inputs required by this model are:

1. The list of demand sites;
2. The list of potential facility sites;

3. The coverage distance (the indicator variable); and
4. Either the *operating cost* of a facility (for the *standardised* set covering model) or, the *fixed location cost* of a new facility (for the *location* set covering model).

The model generates the following outputs:

1. The number of facilities to locate;
2. The location of each facility; and
3. The facility responsible for serving each demand node.

Mathematical notation: *Standard* set covering model (Farahani *et al.*, 2012).

Let

$\mathcal{I}$: set of demand nodes, $\mathcal{I} = \{1, 2, \ldots, i, \ldots, I\}$
$\mathcal{J}$: set of facility sites, $\mathcal{J} \in \{1, 2, \ldots, j, \ldots, J\}$
f_j: the fixed cost of locating a facility at site j
S: the maximum acceptable service distance
a_{ij}: a binary parameter, $a_{ij} = 1$ if the distance from candidate site j to the demand node i is not greater than S.

Decision variable

$$x_j = \begin{cases} 1 & \text{if a facility is to be located at candidate site } j \\ 0 & \text{otherwise.} \end{cases}$$

The task now becomes to minimise

$$\sum_{j=1}^{J} c_j x_j \tag{4.1}$$

subject to

$$\sum_{j=1}^{J} a_{ij} x_j \geq 1 \qquad i \in \mathcal{I}, \tag{4.2}$$

$$x_j \in \{0, 1\} \qquad j \in \mathcal{J}.$$

Mathematical notation: *Location* set covering model (Daskin & Dean, 2004). In the location set covering model, the objective function is to minimise

$$\sum_{j=1}^{J} x_j, \tag{4.3}$$

the constraints in (4.2) apply.

According to Daskin & Dean (2004), a typical difficulty that results from applying the standard set covering model is that the cost of covering all of the demand is often prohibitive. Similarly, when the location set covering model is applied, the number of facilities that need to be located in order to cover all of the demand is often too large.

4.1.2.2 Maximal covering model

The maximal covering location model differs from the set covering model in that it does not necessarily cover all of the demand in the system. It is often used to locate service facilities (such as parks, hospital and schools) and is particularly useful in cases where the budget is insufficcient to cover all of the demand in the network (Davari *et al.*, 2011). The model incorporates data on the demand at each node in order to prioritise the nodes that generate the largest demand. Like the set covering model, the maximal covering model incorporates the concept of coverage distance that dictates the maximum distance between a demand site and the facility that is assigned to serve it.

This model is based on the following assumptions:

1. Facilities can only be located on the nodes of the network;
2. The demand is located at the nodes of the network;
3. The facility capacities are unlimited;
4. The demand nodes are weighted;
5. The coverage distance is fixed;
6. All of the demand is not necessarily covered; and
7. The objective is to maximise the total amount of demand that is covered.

The model requires the following inputs:

1. The list of demand sites;
2. The demand at each site;
3. The list of potential facility sites;
4. The coverage distance (the indicator variable); and
5. The number of facilities to locate.

The outputs generated by the model are:

1. The location of each facility; and
2. The facility responsible for serving the demand nodes that are covered.

Mathematical notation: (Daskin & Dean, 2004).

The definitions for i, j, f_j, S, a_{ij} and x_j are taken from the mathematical notation for the standard set covering problem given in Subsection 4.1.2.1.

Let

h_i: demand at node i

P: the number of facilities to locate.

The decision variable is

$$z_i = \begin{cases} 1 & \text{if demand node } i \text{ is covered} \\ 0 & \text{otherwise.} \end{cases}$$

The objective is now to maximise

$$\sum_{i=1}^{I} h_i z_i \tag{4.4}$$

subject to

$$z_i - \sum_{j=1}^{J} a_{ij} x_j \leq 0 \qquad i \in \mathcal{I}, \tag{4.5}$$

$$\sum_{j=1}^{J} x_j = P, \tag{4.6}$$

$$x_j \in \{0,1\} \qquad j \in \mathcal{J},$$

$$z_i \in \{0,1\} \qquad i \in \mathcal{I}.$$

4.1.2.3 P-center model

The P-center model is similar to the set covering model in that it requires that all of the demand be covered. It is similar to the maximal covering model in that the total number of facilities that are to be located is fixed as part of the constraints. The P-center model manages to cover all of the demand with a fixed number of facilites by relaxing the service standard (i.e. the coverage distance of each facility).

The assumptions of the P-center model are:

1. Facilities can only be located on the nodes of the network;
2. The demand is located at the nodes of the network;
3. The facility capacities are unlimited;
4. The demand nodes are unweighted;
5. The coverage distance is not fixed;
6. All of the demand must be covered; and
7. The goal is to relax the coverage distance only as much as is required to cover all of the demand with the given number of facilities.

The following inputs are required by the model:

1. The list of demand sites;
2. The list of potential facility sites;
3. The number of facilities to locate; and
4. The inter-site travel distances.

The outputs generated by the model are:

1. The coverage distance that was achieved;
2. The location of each facility; and
3. The facility responsible for serving each demand node.

Mathematical notation: (Owen & Daskin, 1998).

The definitions for the variables i, j and P are taken from the mathematical notation for the standard set covering problem and the maximal covering problem given previously.

Let

D: the maximum distance between a demand node and the closest facility

d_{ij}: the distance between demand node i and potential facility site j.

The problem contains two decision variables, the definition for x_j was given in the mathematical notation of the standard set covering problem, the other decision variable is

$$y_{ij} = \begin{cases} 1 & \text{if demand node } i \text{ is served by a facility at node } j \\ 0 & \text{otherwise.} \end{cases}$$

The objective is to minimise

$$D \tag{4.7}$$

subject to

$$\sum_{j=1}^{J} x_j = P, \tag{4.8}$$

$$\sum_{j=1}^{J} y_{ij} = 1 \qquad i \in \mathcal{I}, \tag{4.9}$$

$$y_{ij} - x_j \leq 0 \qquad i \in \mathcal{I},\ j \in \mathcal{J}, \tag{4.10}$$

$$D \geq \sum_{j=1}^{J} d_{ij} y_{ij} \qquad i \in \mathcal{I}, \tag{4.11}$$

$$x_j \in \{0,1\} \qquad j \in \mathcal{J},$$

$$y_{ij} \in \{0,1\} \qquad i \in \mathcal{I},\ j \in \mathcal{J}.$$

4.1.2.4 P-median model

The P-median model is similar to the P-center model in that it also locates a fixed number of facilities in such a way that all of the demand in the model is covered. However, like the maximal covering model, it also takes the demand at each node into consideration. In addition, the P-median model takes the travel distance between each demand node and each candidate facility location into consideration.

The P-median model assumes that:

1. Facilities can only be located on the nodes of the network;
2. The demand is located at the nodes of the network;
3. The facility capacities are unlimited;
4. The demand nodes are weighted;
5. The coverage distance is not fixed;
6. All of the demand must be covered; and
7. The objective is to minimise the total demand-weighted travel distance between each demand node and the facility to which it is assigned.

The inputs required by the P-median model are:

1. The list of demand sites;
2. The demand at each site;
3. The list of potential facility sites;
4. The number of facilities to locate; and
5. The inter-site travel distances.

The model generates the following outputs:

1. The location of each facility; and
2. The facility responsible for serving each demand node.

Mathematical notation: (Owen & Daskin, 1998).

The definitions for the variables i, j, h_i, d_{ij} and P as well as for the decision variables x_j and y_{ij} are taken from the mathematical notation for the standard set covering problem, the maximal covering problem and the P-center model given in Subsections 4.1.2.1 to 4.1.2.3.

The objective is to minimise

$$\sum_{i=1}^{I}\sum_{j=1}^{J} h_i d_{ij} y_{ij} \tag{4.12}$$

subject to

$$\sum_{j=1}^{J} x_j = P, \tag{4.13}$$

$$\sum_{j=1}^{J} y_{ij} = 1 \qquad i \in \mathcal{I}, \tag{4.14}$$

$$y_{ij} - x_j \leq 0 \qquad i \in \mathcal{I},\ j \in \mathcal{J}, \tag{4.15}$$

$$x_j \in \{0,1\} \qquad j \in \mathcal{J},$$

$$y_{ij} \in \{0,1\} \qquad i \in \mathcal{I},\ j \in \mathcal{J}.$$

4.1.2.5 The fixed charge location model

The fixed charge location problem differs from the P-median problem in only two ways: (i) the number of facilities that are to be located are not fixed; and (ii) the objective function is adapted to include the total cost of locating all of the selected facilities. The model is commonly formulated in one of two ways:

1. The *capacitated* fixed charge location model assumes that each facility can serve a limited demand; and
2. The *uncapacitated* fixed charge location model assumes that each facility can serve an unlimited demand.

The assumptions in this model are:

1. Facilities can only be located on the nodes of the network;
2. The demand is located at the nodes of the network;
3. The facility capacities are either limited (*capacitated* model) or unlimited (*uncapacitated* model);
4. The demand nodes are weighted;
5. The coverage distance is not fixed;
6. All of the demand must be covered; and
7. The objective is to minimise the total combined travel cost and fixed facility location cost.

The model requires the following inputs:

1. The list of demand sites;
2. The demand at each site;
3. The list of potential facility sites;
4. The fixed cost of locating a new facility;
5. The inter-site travel distances; and
6. The maximum capacity of each facility (only applicable to the *capacitated* model).

The outputs generated by the model are:

1. The number of facilities to locate;
2. The location of each facility; and
3. The facility responsible for serving each demand node.

Mathematical notation: *Capacitated* fixed charge location model (Current *et al.*, 2002).

The definitions for the variables i, j, f_j, h_i, and d_{ij} are taken from the mathematical notation for the standard set covering problem, the maximal covering problem and the P-center model given previously.

Let

C_j: the capacity of a facility at candidate site j

ω: the transport cost per unit demand and per unit distance.

The decision variables x_j and y_{ij} were defined in Subsections 4.1.2.1 and 4.1.2.3.

The objective now becomes to minimise

$$\sum_{j=1}^{J} f_j x_j + \omega \sum_{i=1}^{I} \sum_{j=1}^{J} h_i d_{ij} y_{ij} \tag{4.16}$$

subject to

$$\sum_{j=1}^{J} y_{ij} = 1 \qquad i \in \mathcal{I}, \tag{4.17}$$

$$y_{ij} - x_j \leq 0 \qquad i \in \mathcal{I},\ j \in \mathcal{J}, \tag{4.18}$$

$$\sum_{j=1}^{J} h_i y_{ij} - C_j x_j \leq 0 \qquad i \in \mathcal{I}, \tag{4.19}$$

$$x_j \in \{0,1\} \qquad j \in \mathcal{J},$$

$$y_{ij} \in \{0,1\} \qquad i \in \mathcal{I},\ j \in \mathcal{J}.$$

Mathematical notation: *Uncapacitated* fixed charge location model (ReVelle & Eiselt, 2005).

The mathematical notation for the *uncapacitated* fixed charge location model is identical to the notation for the *capacitated* fixed charge location model given above, except that constraint (4.19) is omitted.

4.1.2.6 The undesirable facility location (maxisum) model

The maxisum model differs from the P-median in only one way: the model is concerned with locating undesirable facilities (e.g. prisons or solid waste repositories), therefore the model seeks to *maximise* (rather than *minimise*) the total demand-weighted travel distance in the model (Current *et al.*, 2002).

This model is based on the following assumptions:

1. Facilities can only be located on the nodes of the network;
2. The demand is located at the nodes of the network;
3. The facility capacities are unlimited;
4. The demand nodes are weighted;
5. The coverage distance is not fixed;
6. All of the demand must be covered; and
7. The objective is to maximise the total demand-weighted travel distance between each demand node and the facility to which it is assigned.

The inputs required by the model are:

1. The list of demand sites;
2. The demand at each site;
3. The list of potential facility sites;
4. The number of facilities to locate; and
5. The inter-site travel distances.

The outputs generated by the maxisum model are:

1. The location of each facility; and
2. The facility responsible for serving each demand node.

Mathematical notation: (Farahani & Masoud, 2009).

The definitions for the variables i, j, h_i, P and d_{ij} as well as the for the decision variables x_j and y_{ij} are taken from the mathematical notation for the standard set covering problem, the maximal covering problem, and the P-center model given in Subsections 4.1.2.1 to 4.1.2.3.

The objective is to maximise

$$\sum_{i=1}^{I}\sum_{j=1}^{J} h_i d_{ij} y_{ij} \tag{4.20}$$

subject to

$$\sum_{j=1}^{J} y_{ij} = 1 \qquad i \in \mathcal{I}, \tag{4.21}$$

$$\sum_{j=1}^{J} x_j = P, \tag{4.22}$$

$$y_{ij} - x_j \leq 0 \qquad i \in \mathcal{I},\ j \in \mathcal{J}, \tag{4.23}$$

$$x_j \in \{0,1\} \qquad j \in \mathcal{J},$$

$$y_{ij} \in \{0,1\} \qquad i \in \mathcal{I},\ j \in \mathcal{J}.$$

4.1.2.7 The P-dispersion problem

The P-dispersion problem differs from the problems described thus far in two ways: (i) it is not concerned with the distance between the demand sites and the facilities, rather, it is concerned exclusively with the distance between each of the newly sited facilities; and (ii) the model is not concerned with siting facilities in order to serve a defined *demand* that exists within the network, it is merely concerned with siting new facilities. Similar to the maxisum model, the P-dispersion model attempts to *maximise* distance rather than to *minimise* it.

The P-dispersion model assumes that:

1. Facilities can only be located on the nodes of the network;
2. No demand nodes are defined in the network;
3. The facility capacities are irrelevant (and can therefore be viewed as unlimited);
4. No coverage distance exists as there are no demand nodes to cover; and
5. The objective is to site new facilities that are as dispersed as possible in the available space.

The following inputs to the model are required:

1. The number of facilities to locate; and
2. The inter-site travel distances.

The model output is:

1. The location of each facility.

Mathematical notation: (Farahani & Masoud, 2009).

The defnitions for the variables i, j, P and d_{ij} are taken from the mathematical notation for the standard set covering problem, the maximal covering problem, and the P-center model given in Subsections 4.1.2.1 to 4.1.2.3.

Let

E: the minimum distance between facilities

M: a very large number, larger than the maximum d_{ij} value.

The definition for the decision variable x_j was given previously in the mathematical notation of the standard set covering problem.

The objective is to maximise

$$E \tag{4.24}$$

subject to

$$\sum_{j=1}^{J} x_j = P, \tag{4.25}$$

$$E \leq d_{ij} + (M - d_{ij})(1 - x_i) + (M - d_{ij})(1 - x_j) \qquad i \in \mathcal{I},\ i<j, \tag{4.26}$$

$$x_j \in \{0, 1\} \qquad j \in \mathcal{J}.$$

4.1.2.8 The P-hub location problem

Hub location problems are used in designing logistics networks. They differ significantly from all of the models described thus far. The following terminology is used:

1. "Nodes" refers to all of the discrete points that make up the network (similar to the towns, cities and road intersections that exist on a map). In a hub location problem, the "demand" that exists at certain nodes in the network refers to the number of units of material that need to be transported from that node to a specific destination node;

2. "Arcs" refers to the routes connecting all of the nodes in the network (similar to the roads that connect the towns, cities and intersections on the map);

3. The "network" is formed by the system of nodes and arcs (similar to the manner in which the towns, cities, intersections and roads that connect them on a map form a network). There is a cost associated with transporting a unit of material along each arc in the network; and

4. "Hubs" can be viewed as the equivalent of the "facilities" that are placed by every other basic location model that has been discussed thus far and represent transport hubs. A hub is therefore a special kind of node. A key concept underpinning the P-hub location model is that inter-hub transport is more cost-effective than normal inter-node transport or transport between a normal node and a hub.

The decisions that need to be made include where to locate hubs and how to route the deliveries through the network.

The model is based on the following assumptions:

1. Facilities (hubs) can only be located on the nodes of the network;
2. The "demand" is located at the nodes of the network (i.e. the origin nodes);
3. The facility (hub) capacities are unlimited;
4. The arc capacities are unlimited;
5. All of the demand must be covered; and
6. The objective is to minimise the total cost of transporting all of the units of material in the network from their respective origin nodes through the network to their respective destination nodes.

The inputs required by the model are:

1. The list of "demand" sites (origin nodes);
2. The "demand" at each site;
3. The list of potential facility (hub) sites;
4. The number of hubs to locate;
5. The transport cost per unit of material for every arc in the network;
6. The inter-site travel distances; and
7. The discount factor for transport between hubs.

The model generates the following outputs:

1. The location of each facility (hub); and
2. The delivery route for all of the demand in the network.

Mathematical notation: (Current *et al.*, 2002).

The definitions for the variables i, j and P are taken from the mathematical notation for the standard set covering problem and the maximal covering model given in Subsections 4.1.2.1 to 4.1.2.2.

Let

g_{ij}: the number of units of flow between nodes i and j

γ_{ij}: the transportation cost per unit of demand between nodes i and j

β: the discount factor for transportation between hubs.

The definitions of the decision variables x_j and y_{ij} (given in Subsections 4.1.2.1 and 4.1.2.3 respectively) are adjusted only slightly so that they denote the location of a *hub* rather than the location of a *facility*

$$x_j = \begin{cases} 1 & \text{if a hub is to be located at candidate site } j \\ 0 & \text{otherwise,} \end{cases}$$

$$y_{ij} = \begin{cases} 1 & \text{if demand node } i \text{ are assigned to a hub at node } j \\ 0 & \text{otherwise.} \end{cases}$$

The objective is to minimise

$$\sum_{i=1}^{N}\sum_{j=1}^{N} g_{ij}\left(\sum_{k=1}^{N}\gamma_{ik}y_{ik} + \sum_{l=1}^{N}\gamma_{jl}y_{jl} + \sum_{k=1}^{N}\sum_{l=1}^{N}\gamma_{kl}y_{ik}y_{jl}\right) \tag{4.27}$$

subject to

$$\sum_{j=1}^{N} y_{ij} = 1 \qquad i \in \mathcal{I}, \tag{4.28}$$

$$\sum_{j=1}^{J} x_j = P, \tag{4.29}$$

$$y_{ij} - x_j \leq 0 \qquad i \in \mathcal{I},\ j \in \mathcal{J}, \tag{4.30}$$

$$x_j \in \{0,1\} \qquad j \in \mathcal{J},$$

$$y_{ij} \in \{0,1\} \qquad i \in \mathcal{I},\ j \in \mathcal{J}.$$

4.2 Selection of base model for adaptation

Eight basic model types for location science problems were introduced in the previous section. According to Daskin & Dean (2004), the most common basic model types for healthcare applications are the: (i) set covering problem, (ii) maximal covering model, (iii) P-center model, (iv) P-median model and (v) uncapacitated fixed-charge location model. The concept of discrete models versus continuous models was briefly introduced

in Subsection 3.1.1, Daskin & Dean (2004) state that discrete location models are used more extensively than continuous models in healthcare applications. In discrete location models, it is assumed that the demand in a region can be aggregated to a finite number of discrete points. It is also assumed that facilities can be located at a finite number of discrete points. In contrast to this, continuous models assume that demand is distributed across a region. (Often the assumption is that the demand is distributed evenly across the region although this is not a requirement for continuous location modelling.) Similarly, it is assumed that facilities can be located anywhere in the region. The eight basic model types that have been discussed in this thesis are all discrete.

The details of the real-world problem that are to be investigated in this study were described in Subsection 2.3.4. Section 3.3 described how the real-world problem can be interpreted so that it can be modelled using a location science model structure. No detailed analysis of the available input data for modelling the real-world problem has been performed yet. The mathematical formulation and the input data are mutually dependent upon each other. The mathematical formulation determines the input data requirements whilst the availability of input data determines the possible mathematical formulations that can be applied. In this project, the most suitable mathematical formulation for application to the real-world problem will be selected based on (i) the conceptual investigation of the characteristics of the real-world problem presented in Section 2.3, and (ii) two assumptions regarding characteristics of the input data. The available input data will then be analysed in detail to (i) verify the two assumptions that have been made, and (ii) determine whether the required input data sets are available (refer to Section 5.1). Finally, the selected standardised mathematical formulation will be adapted based on the analysis of the input data (refer to Section 5.2).

None of the eight basic model types described in this thesis has the capability to depict the real-world problem with a satisfactory level of accuracy. However, when one recalls the *application-specific nature* of location science models as well as the description of the difficulties associated with *fitting complex, strategic, real-world problems that often involve multiple objectives into theoretical models*, given in Subsection 3.3.1, this is not surprising.

In the absence of a basic model that can accomodate the real-world problem in its standard formulation, the characteristics of the real-world problem will be compared with those of the eight basic models in order to select the most appropriate mathematical formulation to use as a basis for the development of a customised model to solve the real-world problem.

The maxisum model, P-dispersion model and P-hub location model can be removed from consideration without more detailed analysis:

1. P-dispersion model: The real-world problem is not concerned solely with siting new facilities that are as dispersed from one another as possible and the presence of demand is a key component of the real-world problem. Thus the model structure is fundamentally unsuitable as a mechanism to represent the real-world problem;

2. P-hub location model: The real-world problem is not concerned with the design of a logistical network, it does not contain demand that needs to be transported from origin nodes to destination nodes within a network. Therefore this model structure is unsuitable as a mechanism to represent the real-world problem; and

3. Maxisum model: The real-world problem is not concerned with the location of undesirable facilities. Though the rest of this problem formulation does show some promise as a method for representing the real-world problem, it is identical to the P-median problem that will be considered in the detailed analysis. Thus the maxisum model can be safely removed from consideration.

4.2.1 Assumptions regarding the real-world problem

As stated in the previous subsection, the selection of the most appropriate mathematical formulation will be based on two assumptions regarding the real-world problem. These assumptions will be verified as part of the detailed input data analysis in Chapter 5.

4.2.1.1 Assumption on the relationship between loss to follow-up and travel distance

The concept of loss to follow-up was introduced in Subsection 2.3.2. It is a complex phenomenon and it is reasonable to assume that a large number of factors including the psychology of each individual patient, the conduct of the specific healthcare worker that treats the individual patient, the treatment and other conditions at each clinic, the travel time and costs for each patient to reach their healthcare facility, etc. play a role. However, when all of these factors can be assumed to remain approximately constant, and the only difference is in whether the CD4 count is determined using conventional testing or POC testing, it is reasonable to conclude that it is the reduced time delay between when the CD4 sample is taken and when the result of the CD4 test

is available that causes the loss to follow-up rate to decrease when POC testing rather than conventional testing is used.

One component of the time delay between when a CD4 sample is taken and when the result is available (using laboratory-based testing), is the time that it takes to transport the specimen to the closest CD4 testing laboratory. From the above, one could hypothesise that, *facilities that are located further away from their testing laboratory, have a higher loss to follow-up rate.*

This hypothesis will be investigated during the input data analysis in Chapter 5. For now however, the following assumption will be made: The loss to follow-up at each primary healthcare facility is a function of both the *demand* at the site and the *distance* between that site and the closest laboratory.

4.2.1.2 Assumption on testing cost

A detailed investigation of the available costing data will be presented in Chapter 5. For the time being, the following assumption on testing cost will be made: Economies of scale (in terms of the cost per test performed) exist within the laboratory network and for POC testing. Therefore, it is more cost-effective to test a higher volume of CD4 samples than a lower volume of CD4 samples (whether through POC testing or laboratory-based testing) at any one site.

4.2.2 Selection based on model objective

As discussed in Subsection 2.3.4, the real-world problem requires that a number of scenarios for CD4 testing service delivery be investigated based on their (i) financial impact; and (ii) health impact. A safe assumption is that the country would want to maximise the health benefit whilst minimising the cost.

As discussed in Subsection 2.3.2.2, the use of POC CD4 testing (rather than traditional laboratory-based CD4 testing) is expected to result in a reduced number of HIV positive patients being lost to follow-up before ART initiation. Therefore it is logical to conclude that a scenario that relies entirely on POC CD4 testing (rather than laboratory-based CD4 testing) would result in the largest possible health benefit for the country. However, as mentioned in Subsection 2.3.3.2, POC CD4 testing is more expensive than laboratory-based CD4 testing, therefore it is reasonable to conclude that this same scenario would result in the largest possible cost for CD4 testing provision in the country.

From the above, it is clear that the real-world problem involves two conflicting objectives:

1. **Maximise the health impact:** in terms of the problem being investigated here, the health impact can be defined as the number of additional patients being initiated onto ART (i.e. not being lost to follow-up) due to the use of POC CD4 testing; and

2. **Minimise the cost:** for the problem under investigation, the cost will be defined as the total cost of providing CD4 testing to public healthcare facilities in the country, using either laboratory-based or POC technology, or a combination of the two.

The real-world problem is therefore a multi-objective optimisation problem with two objectives. All of the eight basic model structures that have been investigated are single-objective. Literature does contain an increasing number of articles on multi-objective location science applications. In a comprehensive review, Current *et al.* (1990) identified 45 papers that described multi-objective location science applications. Twenty years later, Farahani *et al.* (2010) conducted a review with a slightly wider scope and found approximately 730 articles describing multi-criteria location problems (multi-objective location problems are defined as a subset of multi-criteria location problems). Both studies defined the types of objectives found in the literature:

1. Cost, including fixed and variable costs (identified by both Current *et al.* (1990) and Farahani *et al.* (2010));

2. Coverage and other demand-related objectives (identified by both Current *et al.* (1990) and Farahani *et al.* (2010));

3. Profit (identified by both Current *et al.* (1990) and Farahani *et al.* (2010));

4. Environmental concerns (identified by both Current *et al.* (1990) and Farahani *et al.* (2010));

5. Service level and effectiveness (identified by Farahani *et al.* (2010)); and

6. Other criteria, e.g. resource accessibility and social or political risks (identified by Farahani *et al.* (2010))).

Current *et al.* (1990) found that, from a structural point of view, the multi-objective problems are similar to "the general population of single-objective location models". Farahani *et al.* (2010) echo this conclusion by stating that bi-objective location problems are essentially an extension of the classic, single-objective location science problems. Against this background, it is valid to proceed with selecting a model for adaptation

to the real-world problem from the set of eight basic, single-objective location science problems introduced earlier in this chapter. Table 4.1 summarises the objective of each of the model structures being considered for application to the real-world problem.

Table 4.1: Summarised basic model objectives.

Set covering problem The *standard* set covering model seeks to minimise the total cost of locating the selected facilities. The *location* set covering model aims to minimise the total number of facilities that are located.
Maximal covering model The objective is to maximise the total amount of demand that is covered.
P-center model The goal is to relax the coverage distance only as much as is required to cover all of the demand with the given number of facilities.
P-median model The objective is to minimise the total demand-weighted travel distance between each demand node and the facility to which it is assigned.
The fixed charge location model The objective is to minimise the total combined travel cost and fixed facility location cost.

The maximal covering model's objective seeks to maximise the demand that is covered. In the real-world problem, it is a requirement that all of the demand must be covered, therefore this model's objective is incompatible with the real-world problem.

Both the P-center model and the P-median model are primarily concerned with minimising the distance between the demand nodes and the facilities. The maximum health impact is achieved through the use of POC technology (where the facility is sited at the demand node itself), therefore, minimising the distance between the demand node and the facility would maximise the health impact. These objectives have the potential to be adapted to represent the health impact objective. Neither of the models include a cost consideration.

The two versions of the set covering problem seek to minimise the total cost by minimising either the operating cost or the fixed facility cost. Both of these objectives show potential for being developed into the cost objective to represent the real-world problem.

The fixed charge location model seeks to minimise the combined travel and fixed facility cost. Minimising the travel cost, would seem equivalent to minimising the distance between the demand and the facilities. If this portion of the objective function could be isolated and set as a separate objective, it would have the potential to be adapted to represent the health impact (similar to the suggestion made for the P-center and P-median model). This would then leave the minimisation of the fixed facility

cost as a separate objective. Similar to the discussion of the set covering problem objective, this objective shows potential for representing the cost objective of the real-world problem.

4.2.3 Selection based on model assumptions

In the description of the eight basic models in Subsection 4.1.2, the key assumptions fo each model as well as the input data that is required and the output data that is produced, were summarised. Table 4.2 summarises the assumptions for the five models under consideration.

Table 4.2: Summarised basic model assumptions.

	Set covering problem	**Maximal covering model**	**P-center model**	**P-median model**	**Fixed charge location model**
Facility location	At nodes	At nodes	At nodes	At nodes	At nodes
Demand location	At nodes	At nodes	At nodes	At nodes	At nodes
Facility capacities	Unlimited	Unlimited	Unlimited	Unlimited	Limited[1] or Unlimited[2]
Demand node weighting	Unweighted	Weighted	Unweighted	Weighted	Weighted
Coverage distance	Fixed[3]	Fixed	Not fixed	Not fixed	Not fixed
Demand coverage	Complete	Potentially partial	Complete	Complete	Complete

[1] Only applicable to the capacitated model.
[2] Only applicable to the uncapacitated model.
[3] Though the coverage distance is fixed, it is not necessarily standardised across the model.

As discussed in the introduction to Section 4.2, all of the models that are under consideration are discrete. Thus, all of the models assume that facilities can be located only at the nodes of the network and that demand is sited at the nodes of the network. This is compatible with the real-world problem where demand (in the form of CD4 testing requests) is sited only at each of the approximately 3 600 ART initiation facilities and the facilities that are to be sited by the model (in the form of POC CD4 testing capability) can only be sited at one of these ART initiation facilities or at one of the 61 CD4 laboratories (see Subsection 2.3.4).

The real-world problem does not place a limit on *facility capacity*. One single facility could potentially process all of the CD4 tests that are requested in the country. (This may be viewed as a controversial statement, but, from a purely theoretical point of view, it is true.) Thus, with the exception of the capacitated version of the fixed charge

location model, all of the basic model types under consideration have the capability to adequately depict the real-world problem in terms of facility capacities.

The *demand node weighting* assumption refers to whether a model takes the size of the demand at each node into consideration. Both the set covering problem and the P-center model are unweighted, therefore, the models do not give higher priority to nodes with a larger demand. When considering the two objectives of the real-world problem, it becomes apparent that demand nodes must be weighted in order to model the problem. The health impact objective seeks to maximise the number of additional patients being initiated onto ART. Therefore, when calculating the impact of providing POC CD4 testing capability at a patient-serving healthcare facility, it would be necessary to include an indication of the number of HIV positive patients that have not yet been initiated onto ART but are undergoing CD4 testing at this facility. Similarly, when calculating the cost premium for providing CD4 testing at this facility, it would be necessary to have an indication of the number of CD4 tests that would be performed at the facility. In terms of demand node weighting, the set covering problem and the P-center model are incompatible with the real-world problem, whilst the maximal covering problem, the P-median problem and the fixed charge location model are compatible with the real-world problem.

The real-world problem does not stipulate a required *coverage distance*. It is assumed that that the health impact objective would favour solutions that site CD4 testing capability closer to the patient-serving facilities (refer to the assumption regarding the relationship between loss to follow-up and distance given in Subsection 4.2.1.1). However, there is no limit on the maximum distance between a patient-serving facility and the laboratory that processes the CD4 test requests that are generated there. Thus, all of the model types under consideration have the ability to depict the real-world problem in terms of coverage distance.

It is an absolute requirement of the real-world problem that all of the CD4 test requests that are generated must be tested, either via POC technology or laboratory-based technology. The maximal covering model is therefore incompatible with the real-world problem in terms of demand coverage, whilst the other four models are all compatible.

4.2.4 Selection based on model inputs

The model inputs given in Subsection 4.1.2 are summarised in Table 4.3.

The requirements for a *list of potential facility sites* and a *list of demand sites* as inputs are in line with the discrete nature of the models as described under the *facility*

location and *demand location* assumptions in Subsection 4.2.3. Similarly, the *maximum facility capacity* input relates to the *facility capacities* assumption and the *demand* input relates to the *demand node weighting* assumption, also discussed in Subsection 4.2.3.

Table 4.3: Summarised basic model inputs.

	Set covering problem	**Maximal covering model**	**P-center model**	**P-median model**	**Fixed charge location model**
The list of potential facility sites	√	√	√	√	√
The list of demand sites	√	√	√	√	√
The maximum capacity of each facility					√ [1]
The demand at each site		√		√	√
The coverage distance	√	√			
The inter-site travel distances			√	√	√
The number of facilities to locate	√ [2]	√	√	√	
The cost of a facility	√ [3]				√ [4]

[1] Only applicable to the *capacitated* fixed charge location model.

[2] Only applicable to the *standard* set covering model.

[3] Either the *operating cost* of a facility (for the *standardised* set covering model) or, the *fixed location cost* of a facility (for the *location* set covering model).

[4] The *fixed location cost* of a facility.

The *coverage distance* input relates to the assumption (of the same name). The *inter-site travel distance* input replaces the *coverage distance* input for models that incorporate more complexity when dealing with distance. (Models that make use of a coverage distance are, generally, not concerned with optimising the distance between a demand node and a facility beyond the point of ensuring that this distance does not exceed the coverage distance. The P-center model, which requires inter-site travel distances as an input and then proceeds to specify the minimum required coverage distance, is an exception to this statement.) As discussed in Subsection 4.2.3, siting a testing facility closer to a patient-serving facility improves the performance in terms of the health impact objective. Thus, the P-center model, the P-median model and the fixed charge location model that have the capability to take inter-site travel distances

into consideration are more suited to representing the real-world problem.

The maximal covering model, P-center model, P-median model and one version of the set covering problem require that *the number of facilities that are to be located* be specified as part of the model inputs. As described in Subsection 2.3.4, the real-world problem requires the evaluation of a number of scenarios for the provision of CD4 testing capability. These scenarios range from a solution where all CD4 testing is done via POC technology to a solution where all CD4 testing is done via laboratory-based testing, as well as a number of hybrid solutions that rely on both technologies.

There are various ways to approach the generation of such scenarios, one of which would be to specify different numbers of POC CD4 facilities to locate. If the model had a single objective, so that one optimum answer was produced every time the model was solved, this would be a good approach to scenario generation. However, as discussed in Subsection 4.2.2, the model is multi-objective. According to the description in Subsection 3.1.1, a range of solutions (that form the Pareto set) are produced when a multi-objective model is solved. When one considers the two opposing objectives in the real-world problem (health impact and cost), it becomes clear that the various solutions in the Pareto set will be differentiated from one another by the extent to which POC testing is used. (The solution that uses POC testing the most extensively will have the best performance in terms of health impact and the worst performance in terms of cost, and vice versa.) From the above, it can be concluded that, model formulations that determine the number of facilities that are to be placed endogenously (i.e. the fixed charge location model and the *location* set covering problem), are more suited to represent the real-world problem.

Lastly, it is necessary to take the cost of CD4 testing provision into consideration when modelling the real-world problem. The set covering model and the fixed charge location model have mechanisms for incorporating either the fixed cost or the operating cost of a facility into the problem formulation. The maximal covering model, P-center model and P-median model are incompatible with the real-world problem in terms of this requirement.

4.2.5 Selection based on model outputs

The model outputs for each of the five models (first introduced in Subsection 4.1.2) are summarised in Table 4.4. Subsection 4.2.3 described why *coverage distance* is not a necessary consideration when formulating a model to represent the real word problem. The discussion in Subsection 4.2.4 motivated why a model that determines the *number of facilities* to locate (rather than requiring this information as a model input) is more

suited to modelling the multi-objective real-world problem. It is essential that the model that is used to represent the real-world problem stipulate both the *location of each facility* as well as the *referral pattern* (i.e. the facility responsible for serving each demand site) as part of the model outputs.

Table 4.4: Summarised basic model outputs.

	Set covering problem	**Maximal covering model**	**P-center model**	**P-median model**	**Fixed charge location model**
The coverage distance that was achieved			√		
The number of facilities to locate	√ [1]				√
The location of each facility	√	√	√	√	√
The referral pattern	√	√	√	√	√

[1] Only applicable to the *location* set covering model.

4.2.6 Conclusion: model selection

The five models under consideration have been evaluated for their compatibility with the real-world problem in terms of four categories: (i) objectives; (ii) assumptions; (iii) inputs; and (iv) outputs. The evaluation is summarised in Table 4.5. For each criterium that was considered, a value of "1" indicates that the model is compatible with the real-world problem, a "-1" indicates that the model is incompatible with the real-world problem, and a "0" indicates a neutral impact on compatibility. Each model is given an overall score.

From the discussion in Subsections 4.2.2 to 4.2.5, as well as from the summarised compatibility evaluation in Table 4.5, it is clear that the uncapacitated fixed charge location model (UFL) is most suited to addressing the problem at hand. The UFL determines three key outputs: (i) the number of facilities to locate; (ii) the optimum location of the facilities; and (iii) the optimum referral patterns. All three of these need to be determined as part of the solution to the real-world problem. The UFL covers all of the demand in the system – this is a key requirement of the real-world problem. The UFL takes inter-site distances and costs into consideration and prioritises sites with higher demand, all of these capabilities are required for solving the real-world problem.

Table 4.5: Summarised basic model compatibility evaluation.

	Standard set covering problem	Location set covering problem	Maximal covering model	P-center model	P-median model	Capacitated fixed charge location model	Uncapacitated fixed charge location model
Model objective							
As set out in Subsection 4.1.2	1	1	-1	1	1	1	1
Model assumptions							
Facility location	1	1	1	1	1	1	1
Demand location	1	1	1	1	1	1	1
Facility capacities	1	1	1	1	1	-1	1
Demand node weighting	-1	-1	1	-1	1	1	1
Coverage distance	0	0	0	0	0	0	0
Demand coverage	1	1	-1	1	1	1	1
Model inputs							
List of potential facility sites[1]	0	0	0	0	0	0	0
List of demand sites[2]	0	0	0	0	0	0	0
Maximum facility capacity[3]	0	0	0	0	0	0	0
Demand at each site[4]	0	0	0	0	0	0	0
Coverage distance	0	0	0	0	0	0	0
Inter-site travel distances	-1	-1	-1	1	1	1	1
Number of facilities to locate	-1	1	-1	-1	-1	1	1
Cost of a facility	1	1	-1	-1	-1	1	1
Model outputs							
Coverage distance	0	0	0	0	0	0	0
Number of facilities to locate[5]	0	0	0	0	0	0	0
Location of each facility	1	1	1	1	1	1	1
Referral pattern	1	1	1	1	1	1	1
Total compatibility score	5	7	1	5	7	9	11

[1] Compatibility in terms of this criteria was already evaluated under the "facility location" category in the list of model assumptions.
[2] Compatibility in terms of this criteria was already evaluated under the "demand location" category in the list of model assumptions.
[3] Compatibility in terms of this criteria was already evaluated under the "facility capacities" category in the list of model assumptions.
[4] Compatibility in terms of this criteria was already evaluated under the "demand node weighting" category in the list of model assumptions.
[5] Compatibility in terms of this criteria was already evaluated under the "number of facilities to locate" category in the list of model inputs.

4.3 Conclusion: Mathematical models

This chapter introduced different standard mathematical formulations of location problems and compared the characteristics of these standard formulations with the real-world problem to identify the most suitable model for application in this study.

Although the standard formulation of the UFL model was found to be the most suitable of the eight formulations considered, it is still not directly suited for application to the real-world problem, for the following reasons:

1. The case study has certain unique characteristics which would cause the optimisation mechanism to not function as intended if the standard formulation of the UFL model were applied;

2. The standard formulation of the UFL model does not have the capability to consider all of the factors which need to be considered in order to adequately model the real-world problem; and

3. The formulation of the real-world problem is constrained by the availablity of public healthcare data.

The UFL thus requires modification for this study, and Chapter 5 will describe how the UFL model has been adjusted and applied to the real-world problem.

Chapter 5

Modelling the real-world problem

In the previous chapter, the basic mathematical model that is most suited to the real-world problem was selected based on the general characteristics of the real-world problem. This chapter will describe the modelling and solution of the real-world problem. The following topics will be discussed:

1. Input data analysis;
2. Tailored mathematical model;
3. Model validation;
4. Solution methodology; and
5. Solution methodology verification.

5.1 Input data

In Subsection 3.3.1, reference was made to the difficulties associated with applying the theoretical location science models that have been developed to solve real-world problems. In the South African public healthcare context under investigation, the largest challenge associated with modelling the real-world problem has been the availability of relevant, representative and accurate data.

Two key assumptions regarding the real-world problem were made when selecting the most appropriate mathematical formulation (Subsection 4.2.1). These assumptions need to be verified by analysing the available input data. This section starts with an explanation of the role these two assumptions play in the functioning of the optimisation mechanism of the proposed model formulation.

The input data requirements of the real-world problem were introduced in Subsection 4.2.4. The remainder of this section discusses how each of the input data sets were formulated by describing the raw data sets that were sourced as well as how these data sets were processed and analysed.

5.1.1 Input data assumption verification

Two assumptions regarding the real-world problem were used when selecting the most appropriate mathematical formulation. These assumptions were stated in Subsections 4.2.1.1 and 4.2.1.2, and are repeated here:

1. The loss to follow-up at each primary healthcare site is a function of both the *demand* at the site and the *distance* between that site and the closest laboratory; and

2. Economies of scale (in terms of the cost per test performed) exist within the laboratory network and for POC testing. Therefore, it is more cost-effective to test a higher volume of CD4 samples than a lower volume of CD4 samples (whether through POC testing or laboratory-based testing) at any one site.

These two assumptions will be investigated during the input data analysis in Subsections 5.1.4 and 5.1.7. In order to provide the reader with an understanding of the importance of these assumptions, the remainder of this section will describe: (i) how it is proposed that the standard UFL model's objective should be transformed in order to model the real-world problem; and (ii) what role each of these assumptions fulfill in the optimisation mechanism of the transformed model objectives.

5.1.1.1 The need for two objectives

The most important requirement of the case study not addressed by the standard formulation of the UFL is the ability to consider the health impact as part of the model objective. In the case study, an improved health impact would necessarily cost more. Therefore, an improvement in health impact weakens the performance in terms of the standard model's primary (and only objective), cost. Due to this relationship between health impact and cost, combining the cost and the inverse of the health impact into a single objective that needs to be minimised will not produce a satisfactory result. Therefore the standard UFL model is adapted by:

1. Transforming it to a multi-objective optimisation (MOO) model; and

2. Setting the second objective as the health impact.

5.1.1.2 Proposed changes to the cost objective

In the standard formulation, the primary (and only) objective is cost. This objective has two components: (i) the total cost of locating the selected facilities; and (ii) the total demand-weighted travel time. The cost objective is adjusted in two ways in order to make it suitable for application to the study:

1. The location cost component is transformed to the total testing cost (for the combination of conventional testing and POC testing that is used) for one year. This total testing cost includes the reagents and consumables that are used, as well as the equipment costs and the labour costs; and

2. The demand-weighted travel component is removed from the cost calculation. In the real-world problem being investigated, the allocation of POC CD4 testing capability to a patient-serving healthcare facility has no impact on the total amount of transport required to deliver the diagnostic service in the country. This is because the specimen transport service of the NHLS collects specimens for a wide variety of diagnostic tests, including CD4, in a single collection round. Removing CD4 testing from the batch of specimens that need to be transported does not reduce the urgency or the required collection frequency of the batch of specimens. The inclusion of both the demand and either the travel distance or travel time in the model's objective(s) is key in ensuring that both the location of the facilities and the referral pattern is optimised. The possibility of using an arbitrary weighting factor for the "cost" of the demand-weigthed travel time and still including this in the cost objective was considered. If this approach were followed, the ratio between the travel costs and the total testing costs could be adjusted by adjusting the weighting factor. However, this approach is not ideal as the model would show great sensitivity to the ratio between the travel costs and the total testing costs and this ratio would be arbitrary.

In its adjusted form, the cost objective must perform the following function within the optimisation mechanism: it must ensure that, for a given health impact, an optimal number of POC facilities and laboratory facilities is located. This can be achieved if one can assume that *economies of scale exist* (i.e. that it is more cost-effective to test a higher volume of samples at a site). This assumption was introduced in Subsection 4.2.1.2. It was repeated in Subsection 5.1.1, and, as mentioned, it will be evaluated by analysing the available costing data in Subsection 5.1.4.

5.1.1.3 The proposed health impact objective

Health impact (the second objective) is defined as the number of additional people that are initiated onto ART in a year due to the changes that have been made to the diagnostic service provision model.

The following hypothesis relating to the relationship between loss to follow-up and distance was introduced in Subsection 4.2.1.1: *facilities that are located further away from their testing laboratory, have a higher loss to follow-up rate.*

If this hypothesis is correct, the health impact objective will serve to optimise the referral patterns by minimising the total demand-weighted distance between the primary healthcare facilities and the laboratories that they are assigned to. Furthermore, if the hypothesis is correct, then the assumption regarding the relationship between the loss to follow-up at each primary healthcare facility and the *demand* at the site as well as the *distance* between that site and the closest laboratory (introduced in Subsection 4.2.1.2 and repeated in Subsection 5.1.1) would be proven valid.

From the above, the calculation of the health impact objective will have two components:

1. For primary healthcare facilities that are assigned to laboratory-based testing, the facility's loss to follow-up rate will be determined based on the distance between that facility and the CD4 laboratory it is assigned to.

2. For facilities that are assigned to POC testing, the facility's loss to follow-up will be determined by applying the odds ratio associated with POC testing (the Wynberg *et al.* (2014) odds ratio introduced in Subsection 2.3.2.) to the laboratory-based loss to follow-up rate determined based on the distance between the facility and the closest existing CD4 laboratory.

5.1.2 Input data set: List of primary healthcare facilities offering ART initiation

The raw data sets numbers two and three listed in Table 5.1 were used to determine the list of primary healthcare facilities that offer ART initiation to be included in the model. All primary healthcare facilities that offer ART initiation should report ART statistics via the District Health Information System (DHIS). The DHIS data set that contains the ART initiation statistics does, however, not contain any facility co-ordinates. There is no unique identifier that can be used to match the records in the two DHIS data sets that were used, however, the facility naming convention used in the

Table 5.1: Raw input data sets.

1	Master list of DOH facilities	DHIS[1]
2	List of DOH facility co-ordinates	DHIS[2]
3	ART statistics reported during the 2013/2014 financial year	DHIS[3]
4	CD4 testing volumes for the financial year 2012/2013	NHLS[4]
5	List of NHLS CD4 laboratories and their co-ordinates	NHLS[6]
6	List of NHLS non-CD4 laboratories and their co-ordinates	NHLS[5]
7	Post-ART initiation LTFU data	tier.net[7]

[1] Via a data request submitted to the National Health Laboratory Service on 14 June 2013. Data supplied via email on 14 June 2013.

[2] Via a data request submitted to the Department of Health on 24 April 2014. Data received via email on 8 May 2014.

[3] Data set provided by the Clinton Health Access Initiative, received via email on 2 May 2014.

[4] Via a data request submitted to the National Health Laboratory Service on 4 March 2014. Data supplied via email on 10 April 2014.

[5] Via a data request submitted to the National Health Laboratory Service on 6 May 2014. Data supplied via email on 6 May 2014.

[6] Via a data request submitted to the National Health Laboratory Service on 14 May 2014. Data supplied via email on 16 May 2014.

[7] Data set provided by the Clinton Health Access Initiative, received via email on 16 May 2014.

two data sets is similar. Co-ordinate data was matched to the list of facilities offering ART initiation, based on the facility name. According to the DHIS ART data set, 3 633 primary healthcare facilities were offering ART initiation in January 2014. DHIS co-ordinate data was available for 3 587 of these facilities. The 46 facilities for which no co-ordinate data is available represent 0.3% of the total number of ART patients in SA (based on the January 2014 data). These 46 sites have been excluded from the study.

5.1.3 Input data set: Demand per primary healthcare facility

Raw input data sets numbers one to four listed in Table 5.1 were used in the calculation of the demand per primary healthcare facility. The following data processing steps were executed:

1. The CD4 testing volumes for the 2012/2013 financial year (data set number four) were matched to the master list of DOH facilities. The data sets did not contain a unique identifier for the facilities that could be used to match the records. The naming convention used in the two data sets did not allow for the use of any automated methods in matching the records, therefore the records were matched manually. It is interesting to note that certain primary healthcare facilities appear to be duplicated up to six times in the NHLS data set. CD4 testing volumes were successfully matched to 4 041 primary healthcare facilities. 4.7% of the total CD4 testing volume could not be matched to a facility.

2. Unique identifiers were used to match the data set above to the set of 3 587 primary healthcare facilities that report ART statistics and for which co-ordinate data is available (refer to Subsection 5.1.2). No CD4 testing data could be matched for 308 of the 3 587 facilities. These 308 facilities account for 7.2% of the total number of newly initiated ART patients in the 2013/2014 financial year. Initially, these facilities were included in the model. However, following the presentation of preliminary results to the NHLS, a decision was made in collaboration with subject matter experts from the NHLS[1] to exclude these sites from the modelling exercise (Stevens *et al.*, 2014).

[1]Prof. Wendy Stevens (*Head of the NHLS' National Priority Programme and Head of Molecular Medicine and Haematology at the University of the Witwatersrand*), Prof. Debbie Glencross (*Director of the NHLS' CD4 National Priority Programme and the Pathologist in Charge of the Charlotte Maxeke Johannesburg Academic Hospital's CD4, Leukaemia, Immunohaematology and HIV Immunology Laboratory*), Dr Lindi Coetzee (*CD4 Co-ordinator of the NHLS' National Priority Programme*), and Naseem Cassim (*NHLS National Priority Programme staff member*).

This data gives the total number of CD4 tests that were performed for determining ART initiation eligibility and for monitoring patients that are receiving ART. There is no reputable data set available that would allow one to distinguish between the tests that have been requested for the purpose of determining eligibility and those that have been requested for the purpose of monitoring ART patients. The fact that this data is not available has been confirmed by the NHLS subject matter experts (Stevens *et al.*, 2014). However, it is the recommendation of various subject matter experts from the Clinton Health Access Initiative[1] (CHAI), WITS University's Health Economics and Epidemiology Research Office[2] (HEERO) and the NHLS that it would not be practically feasible to exercise the required control within clinics and hospitals to ensure that POC CD4 devices are only used for determining eligibility for ART initiation[3] and not for monitoring ART patients[4] (Lehe *et al.*, 2014). For these reasons, the modelling of the real-world problem will include the cost of all CD4 testing in the country (thus the cost of CD4 testing for determining eligibility for ART initiation as well as the cost of CD4 testing for monitoring ART patients.) Furthermore, the model will work on the assumption that, if POC CD4 testing capability is installed at a facility, both the eligibility testing and the monitoring testing will be performed using POC.

5.1.4 Input data set: Cost per test

Costing data was taken from Cassim *et al.* (2014). This study is currently in the peer review process of the Plos ONE journal. The study was conducted using data from the NHLS in South Africa. A tiered service delivery model for the provision of CD4 tests was proposed. The tiers were defined based on the testing method (either laboratory-based or POC) and the volume of samples that are tested per day. The study calculated a cost per test that is inclusive of reagent, consumable, staffing and equipment costs. Test error rates (published rates for POC testing and in-house NHLS rates for laboratory-based testing) were taken into account to ensure that the cost per test provides for tests that would need to be executed more than once. Sample

[1] Jonathan Lehe (*Senior Manager: Point of Care Diagnostics*) and Damian Fuller (*Lesotho Country Support Associate for MDR-TB*).

[2] Kate Schnippel (*Senior Researcher*)

[3] POC CD4 tests offer a positive health impact (in terms of the number of eligible individuals that are initiated onto ART) when used for determining eligibility for ART initiation.

[4] It is possible that using POC CD4 testing for monitoring ART patients may offer positive health impacts. However, for the purpose of this study a positive health impact is defined as reducing the loss to follow-up prior to ART initiation. When POC CD4 testing is used for monitoring ART patients, it has no effect on the loss to follow-up prior to ART initiaiton.

transportation costs, laboratory management costs, facility rental and maintenance costs, etc. were excluded from the study. The five testing tiers, with the type of testing used, the maximum / minimum testing volume and the cost per test is summarised in Table 5.2. This data supports the *economies of scale* assumptions listed in Subsection 5.1.1.

Table 5.2: Cost per test data. (As calculated by Cassim *et al.* (2014).)

Tier	Type of testing	Daily testing volume	Cost per test (US$)[1]
1	POC	≤5	35.46
2	POC	≤40	19.02
3	Conventional	≤100	10.56
4	Conventional	≤394	9.38
5	Conventional	>394	8.51

[1] Because the paper by Cassim *et al.* (2014) is still in the peer review process, this is a hidden cost. However, both the order of magnitude of the values and the ratios between the values are accurate.

5.1.5 List of potential POC testing sites

Raw input data sets numbers one to four listed in Table 5.1 were processed as described in Subsection 5.1.3 to generate the list of 3 279 facilities that were to be included in the final model. In accordance with the maximum daily testing volume for POC sites as defined by Cassim *et al.* (2014), only sites that request a maximum of 9 600 CD4 tests per year were considered eligible for POC testing. (The testing threshold of 9 600 is calculated based on the assumption of 20 working days per month given in the Cassim *et al.* (2014) study.)

5.1.6 Input data set: List of CD4 laboratory sites

Raw input data set number five in Table 5.1 contains a list of the NHLS laboratories that currently offer CD4 testing, together with the facility co-ordinates. This data was not processed in any way.

5.1.7 Analysis of relationship between distance and LTFU

Subsection 4.2.1.1 stated that a key assumption that was made when selecting the most appropriate mathematical formulation for the real-world problem was that the loss to

follow-up rate at each primary healthcare facility is a function of both the *demand* at the site and the *distance* between that site and the closest laboratory. This assumption was examined in Subsection 5.1.1.3 and it is concluded that if the following hypothesis is true, the assumption can be assumed to be valid: *facilities that are located further away from their testing laboratory, have a higher loss to follow-up.* This hypothesis was tested using data sets seven and eight listed in Table 5.1.

As discussed in Subsection 2.3.2.2, the use of POC CD4 testing is expected to impact the *pre* ART initiation loss to follow-up rate. Unfortunately there is no representative data set of *pre* ART initiation loss to follow-up data available for the country. However, at present, 648 public healthcare facilities in the country report *post* ART initiation loss to follow-up statistics (data set seven in Table 5.1). In the absence of representative *pre* ART initiation LTFU data for the country, a decision was made to investigate a relationship between *post* ART initiation LTFU data and the distance between each of the healthcare facilities and the closest CD4 testing laboratory (data set eight from Table 5.1).

A Pearson test and a Spearman rank test were performed and neither test showed evidence of a correlation between the two data sets. The Pearson test result was r=-0.0718, p=0.0764 and the Spearman rank test result was $\rho(622)$=-0.0365, p=0.3000. The data was analysed to determine whether other factors (such as the urban vs. rural classification of a site, or the size of the site in terms of patient numbers) impacted the loss to follow-up rate at the site. Stellenbosch University's Centre for Statistical Consultation was also consulted to verify that the data analysis methodology that was applied was correct. These subject matter experts agreeed that, even when one controlled for specific factors (such as those listed earlier in this paragraph), there was still no evidence of a correlation between the two variables.

From the results of the correlation analysis above, we reject the hypothesis that *facilities that are located further away from their testing laboratory, have a higher loss to follow-up.* From this it follows that the assumption regarding the relationship between the health impact objective and the *demand* at the site as well as the *distance* between that site and the closest laboratory, has not been proven correct.

The proposed formulation of the health impact objective as well as the functions that would be performed by this objective was introduced in Subsection 5.1.1.3. Since no evidence of a relationship between distance and the LTFU rate could be established, the health impact objective must be adjusted so that it no longer takes the distance between a site and its closest laboratory into consideration. Therefore, this objective will no longer serve to minimise the total demand-weighted distance between primary

healthcare facilities and the laboratories to which they refer their work. It is therefore necessary to control this in another way. The use of a coverage distance is proposed, this will be discussed in Subsection 5.1.9.

5.1.8 Health impact of locating a POC facility

In view of the fact that the assumption that the health impact is a function of both the demand at a facility and the distance between that facility and the closest laboratory could not be confirmed as correct, the proposed formulation of the health impact objective must be adjusted.

In Section 5.1 it was stated that the availability of relevant, representative and accurate data has been the largest challenge associated with modelling the real-world. This is particularly applicable to data regarding the likely health impact of different CD4 testing service delivery models.

A large number of studies on the LTFU rate prior to ART initiation (based on laboratory-based CD4 testing) has been conducted. A systematic review by Mugglin *et al.* (2012) was used to introduce the concept of loss to follow-up in Subsection 2.3.2.1. Though Mugglin *et al.* (2012) set out the concept very clearly, a systematic review by Rosen & Fox (2011) has been cited significantly more frequently than Mugglin *et al.* (2012)[1]. For this reason, the Rosen & Fox (2011) study will be used when calculating the health impact of locating a POC facility. More recently, a large number of studies has been conducted to determine the impact of POC CD4 testing on the LTFU rate prior to ART initiation, Wynberg *et al.* (2014) systematically reviewed this literature. The results of these systematic reviews are summarised in Table 5.3. The use of the Rosen & Fox (2011) and Wynberg *et al.* (2014) data in this study was approved by the subject matter experts from HEERO and those from CHAI.

Data set number three listed in Table 5.1 indicates the total number of patients that were newly initiated onto ART at each facility each month (using laboratory-based CD4 testing). This number includes: patients that were eligible for immediate treatment initiation based on their CD4 results, patients that had been enrolled in pre-ART care (based on their CD4 result) before becoming eligible for ART; and patients that were eligible for ART initiation regardless of their CD4 count (e.g. pregnant patients and those with active TB). If it were possible to extract from this data (i) the number of patients that were initiated onto ART immediately (δ); and (ii) the number of patients that were initiated onto ART after having been enrolled in pre-ART care (κ), it would

[1] On 04/10/2014, Scopus reported that Rosen & Fox (2011) had been cited 169 times whilst Mugglin *et al.* (2012) had been cited 10 times.

Table 5.3: Summary of literature findings on pre-ART initiation LTFU rate.

Treatment stage	**LTFU rate based on laboratory CD4 testing**[1]	**Odds ratio based on POC CD4 testing**[2]	**LTFU rate based on POC CD4 testing**[3]
HIV testing to CD4 testing	41%	4.1	14% to 20%[4]
CD4 testing to receipt of result		2.8	
CD4 result receipt to ART initiation (patients that are not eligible for immediate ART initiation, and are therefore enrolled in pre-ART care until they become eligible)	54%	1.8	39%
CD4 result receipt to ART initiation (patients that are eligible for immediate ART initiation, based on their CD4 results)	32%	0.98	32%

[1] According to Rosen & Fox (2011) systematic review.

[2] According to Wynberg *et al.* (2014) systematic review.

[3] Calculated by applying Wynberg *et al.* (2014) odds ratio to Rosen & Fox (2011) LTFU rate.

[4] The Rosen & Fox (2011) review reports a single LTFU rate for the treatment stage from HIV testing to receipt of CD testing result, while the Wynberg *et al.* (2014) review breaks this up into two separate treatment steps and reports a unique odds ratio for each step.

be possible to use the LTFU rates given in Table 5.3 to determine the number of patients that would have been initiated onto treatment if POC testing had been used at each site(ζ) using the following expression

$$\zeta = \left[\frac{1-0.32}{1-0.32}\delta + \frac{1-0.39}{1-0.54}\kappa\right]\left[\frac{1-0.2}{1-0.41}\right]$$
$$\zeta = 1.36\delta + 1.79\kappa.$$

This formula follows a conservative approach, the POC LTFU rate for the stage from HIV testing to receipt of CD4 result has been set to the upper limit of 20%.

Data set number four in Table 5.1 indicates the number of CD4 test results that fell within each of the following bands during the year: $r \leq 50$; $50 < r \leq 100$; $100 < r \leq 200$; $200 < r \leq 350$; $350 < r \leq 500$; $r < 50$. However, this level of detail is not sufficient to allow one to determine what percentage of patients were eligible for treament immediately after their first CD4 test and what percentage were enrolled in pre-ART care first:

1. As stated in Subsection 5.1.3, no data is available on the proportion of the reported CD4 tests that were conducted for the purpose of determining eligibility for ART initiation versus those that were conducted for the purpose of monitoring patients that were already on ART therapy; and

2. A large number of factors influences a patient's CD4 count. Therefore, though one would expect an individual's CD4 count to increase after initiating ART, it is incorrect to assume that all CD4 tests with a result of less than 350 were conducted for the purpose of determining eligibility for ART initiation.

As described, it was not possible to distinguish between the number of patients that were initiated onto ART immediately, those that were enrolled in pre-ART care first, and those that were initiated onto treatment regardless of their CD4 count. Therefore, a decision was taken to apply the most conservative factor (1.36) to the available data at each facility (i.e. the total number of patients initiated onto ART at the facility per year) in order to calculate the likely impact of using POC testing at a specific site.

5.1.9 The need for a coverage distance

Subsection 5.1.7 concluded by proposing that a coverage distance should be incorporated into the model in order to control the distance between the primary healthcare facilties and the CD4 testing laboratories that they refer to. Two recent South African-based studies propose a coverage distance for public healthcare laboratories providing

CD4 testing. Cassim *et al.* (2014) propose that the tier three laboratories (described in Subsection 5.1.4) should serve facilities that lie within a 100km radius[1]. The study also states that, the tier three laboratories in the Pixley ka Seme district in the Northern Cape were serving primary healthcare facilities within a radius of 50km to 250km[2]. Furthermore, Cassim *et al.* (2014) suggests that a distance of more than 150km between the primary healthcare facility and the CD4 testing laboratory may negatively affect both the specimen integrity and the turnaround time of the CD4 test result delivery to the primary healthcare facility. Glencross *et al.* (2014) in their investigation of the establishment of a tiered laboratory network for CD4 service provision, suggested CD4 laboratory coverage radiuses of between 50km and 200km. This study states that facilities that fall outside a 100km radius of a CD4 testing laboratory, "largely coincide with districts with poorer turnaround time (of laboratory testing)". Literature therefore does not agree on a specific acceptable coverage radius for CD4 testing laboratories, although radii of between 50km and 200km are suggested (and radiuses of up to 250km are observed in practice). There also is no clear consensus on the distance at which turnaround performance starts being affected, though both 100km and 150km are proposed. In view of the lack of a clear consensus on coverage distance in literature, a CD4 laboratory coverage distance of 150km is used in the model. Because this distance is not supported by clear evidence, sensitivity analysis is performed on this variable by also testing the impact of setting the coverage distance to 100km and to 200km respectively.

5.1.10 Input data set: Inter-site travel distances

The euclidian distance between each of the ART initiation healthcare facilities and each of the potential CD4 testing sites was calculated based on data sets numbers two, five and six in Table 5.1. As discussed in the previous subsection, a coverage radius will be used in the model, therefore it is appropriate to use the euclidian distance between facilities rather than the actual travel distance via the road network.

[1] The draft paper proposes that these laboratories should serve an area of 100km^2. However when one reads the context in which this statement is made, it is reasonable to conclude that the authors meant to suggest a coverage radius of 100km. This conclusion was verified during a meeting with three of the authors (Mr N. Cassim, Dr L. Coetzee and Prof. D. Glencross) on 10 June 2014.

[2] In the draft article, the unit of this radius is given as "km^2", however, the value is clearly labelled "radius of referrals" therefore this unit can be assumed to be an oversight.

5.1.11 Summarised final input data set

The input data sets that have been included in the final model are summarised in Table 5.4.

Table 5.4: Summarised final input data set.

A list of 3 279[1] ART initiation facilities with the following attributes: The total number of CD4 tests requested by the facility during the 2012/2013 financial year (refer to Subsection 5.1.3). The number of additional patients that would have been initiated onto ART at the facility during the 2013/2014 financial year if POC CD4 testing had been used at the facility (refer to Subsection 5.1.8).
A cost per test for each of the five tiers defined by Cassim *et al.* (2014) (refer to Subsection 5.1.4).
A list of 61 CD4 laboratories (refer to Subsection 5.1.6)
An inter-site travel distance matrix (refer to Subsection 5.1.10)

[1] Refer to Subsections 5.1.2 and 5.1.3 for an explanation of the number of ART initiation facilities that are included in the model.

5.2 Tailored mathematical model

The next four subsections describe how the standard formulation has been adjusted in order to model the real-world problem with the available public healthcare data. The mathematical formulation of the real-world problem is given in Subsection 5.2.5.

5.2.1 Adjustments to the model objectives

The proposed adjustments to the model objectives were discussed in Subsections 5.1.1.1 to 5.1.1.3. The assumption on which the proposed changes to the cost objective were based, has been proven correct, therefore this objective has been adjusted as described in Subsection 5.1.1.2. The assumption on which the proposed changes to the health impact objective were based has not been proven correct, therefore the changes proposed in Subsection 5.1.1.3 will not be implemented. Instead, the health impact objective will be adjusted as described below.

The health impact is defined as the number of *additional* people that are initiated onto ART (over and above the number of patients that are currently initiated, using laboratory-based CD4 testing). Therefore:

1. The calculation of the health impact objective will have only *one* component. For facilities where POC CD4 testing is allocated, the factor derived in Subsection

5.1.8 will be applied to data on the number of patients initiated at the facility during the 2013/2014 financial year to calculate the number of additional patients that will be initiated onto treatment due to the use of POC. Facilities that make use of conventional testing will be assumed to have no incremental health impact; and

2. The health impact objective will *not* serve to minimise the total demand-weighted travel time between healthcare facilities and the CD4 laboratories that they have been assigned to. Instead, as proposed in Subsection 5.1.7, this will be controlled through the use of a coverage distance.

5.2.2 Adjustments to the model assumptions

With the exception of the assumption regarding the need for a coverage distance, all of the assumptions that the UFL is based on, are applicable to the real-world problem. The need for a coverage distance was introduced in Subsection 5.1.7. Please refer to Subsection 4.2.3 for a detailed discussion of the applicability of each of the remaining assumptions.

5.2.3 Adjustments to the model inputs

The availability of input data, the processing of the available data and the summarised final input data set for the real-world problem have been summarised in Section 5.1. This input data set differs from the UFL input data requirements in only two ways: (i) the need for a coverage distance as input to the model, refer to Subsection 5.1.9; and (ii) the nature of the cost data that is required as an input, refer to Subsections 5.1.1.2 and 5.1.4.

5.2.4 Adjustments to the model outputs

The outputs of the UFL model are appropriate for modelling the real-world problem, this was discussed in more detail in Subsection 4.2.5.

5.2.5 Mathematical formulation of real-world problem

The mathematical formulation of the real-world problem given here incorporates the required changes to the standard UFL model's objectives, assumptions and inputs.

Let

$\mathcal{I}$: set of ART initiation facilities, $i \in \{1, 2, \ldots, i, \ldots, I\}$

$\mathcal{J}$: set of CD4 testing facilities, $j \in \{1, 2, \ldots, j, \ldots, J\}$
where $j = 1$ represents POC testing and
$j = 2, 3, \ldots, J$ represent laboratory sites

S: the maximum acceptable distance between an ART initiation facility and the laboratory facility that it is assigned to

n_i: the number of additional patients that would have been initiated onto ART at site i if POC testing had been used at the site

d_{ij}: the distance between ART initiation site i and laboratory facility j

v_i: the number of CD4 tests requested by ART initiation site i

w_j: the total CD4 testing volume assigned to laboratory j

q_j: the cost of performing the CD4 tests assigned to laboratory j

r_i: the cost of performing the POC tests assigned to ART initiation site i.

The decision variable is

$$y_{ij} = \begin{cases} 1 & \text{if the CD4 testing for ART initiation facility } i \text{ is assigned to} \\ & \text{testing facility } j \\ 0 & \text{otherwise.} \end{cases}$$

The objectives are to maximise

$$\sum_{i=1}^{I} y_{i1} n_i \tag{5.1}$$

and minimise

$$\sum_{i=1}^{I} r_i + \sum_{j=2}^{J} q_j \tag{5.2}$$

subject to

$$\sum_{j=1}^{J} y_{ij} = 1 \qquad i \in \mathcal{I}, \tag{5.3}$$

$$y_{ij} d_{ij} \leq S \qquad j \in \mathcal{J},\ i \in \mathcal{I}, \tag{5.4}$$

$$w_j = \sum_{i=1}^{I} y_{ij} v_i \qquad j = 2, 3, \ldots, J, \tag{5.5}$$

$$q_j = \begin{cases} 10,56 w_j & \text{if } w_j \leq 24\,000 \\ 9,38 w_j & \text{if } w_j \leq 94\,560 \\ 8,51 w_j & \text{if } w_j > 94\,560 \end{cases} \qquad j = 2, 3, \ldots, J, \tag{5.6}$$

$$r_i = \begin{cases} 35,46 x_{i1} v_i & \text{if } y_{i1} v_i \leq 1200 \\ 19,02 x_{i1} v_i & \text{if } y_{i1} v_i > 1200 \end{cases} \qquad i \in \mathcal{I}, \tag{5.7}$$

$$y_{i1} v_i \leq 9\,600 \qquad i \in \mathcal{I}, \tag{5.8}$$

y_{ij} is binary.

The coefficients in (5.6) and (5.7) are equivalent to the cost per test data in Table 5.2. The constants in these two equations represent the annual testing volume thresholds for each costing tier. These annual volumes were calculated based on an assumption of 20 working days per month, this is the same assumption that was used by Cassim *et al.* (2014) when defining the daily testing volume thresholds for each tier in Table 5.2. Constraint (5.8) ensures that POC testing is not assigned to sites that process more than the maximum annual testing volume for POC sites as defined by Cassim *et al.* (2014).

5.3 Validation of the problem formulation

The problem formulation has been validated to confirm that it adequately depicts the real-world problem through presentation to three sets of subject matter experts:

1. Public healthcare experts from CHAI and a health economist from HEERO were frequently consulted during the selection of the most appropriate standard mathematical formulation, the analysis of the available input data and the development of the tailored mathematical formulation. These subject matter experts were consulted in order to ensure that the assumptions made were reasonable, that the input data and other literature were interpreted correctly and that the characteristics of the real-world problem were understood correctly;

2. The selected standard mathematical formulation and the input data analysis were presented to the NHLS subject matter experts (Stevens *et al.*, 2014); and

3. The problem formulation and the input data analysis were presented at the triennial conference of the International Federation of Operations Research Societies in Spain (2014).

5.4 Solution methodology

A number of factors contribute to make this a difficult problem to solve:

1. The decision space is extremely large (a binary matrix of 3 279 rows by 62 columns, with only one "1" allowed per row);

2. The problem is non-linear (specifically, the tiered cost per test values given in Table 5.2 create a saw-tooth function); and

3. The problem is multi-objective.

The first approach was to attempt to solve the problem using the commercial package AIMMS. The reasons for selecting AIMMS were: (i) the software has a very user-friendly interface; (ii) AIMMS contains a large variety of built-in algorithms; (iii) the package has been successfully used to solve complex models both by the researcher and by fellow M.Eng students at Stellenbosch University; and (iv) the researcher is familiar with the software. AIMMS does not have the capability to solve multi-objective problems, therefore the problem was converted to a single-objective problem by fixing the available budget and setting this as a constraint, leaving the health impact as the only objective. This model formulation was tested using a small version of the problem (i.e. a reduced number of DOH and laboratory sites). AIMMS was unable to solve this single-objective version of the problem. Further investigation confirmed that this was due to the non-linear, saw-tooth cost function. (This was confirmed by replacing the cost function with a linear function, AIMMS was then able to solve the problem.)

As a second attempt at solving the model, a metaheuristic for multi-objective optimisation was applied to the problem. The Cross Entropy Method for Multi-objective Optimisation (the MOO CEM) developed by Bekker (2012) was selected. The reasons for selecting this metaheuristic were: (i) the MOOCEM has the ability to solve problems with multiple objectives; and (ii) literature describes that the metaheuristic has been successfully applied to similar problems (more details are given in the following subsection). Both the problem formulation and the metaheuristic were adjusted, first

to enable the metaheuristic to be successfully applied to the problem and second to improve the efficiency with which the metaheuristic was able to solve the problem. The application of the MOO CEM to the real-world problem is described in more detail in the next subsection.

In order to verify the quality of the solution generated by the MOO CEM, attempts were made to apply various built-in algorithms in the Matlab software package to the real-world problem. A single-objective version of the problem was succesfully solved with the Matlab Genetic Algorithm. The process of selecting this algorithm and applying it to the real-world problem is described in more detail in Subsection 5.4.2.

5.4.1 Multi-objective solution via the MOO CEM

The cross-entropy method was developed by Rubinstein (1997) and extended for application to single-objective continuous and combinatorial problems by Rubinstein & Kroese (2004). The cross-entropy method has been adapted for application to both *continuous* multi-objective optimisation problems (Bekker & Aldrich, 2011) and to *discrete* multi-objective optimisation problems (Bekker, 2013). Both Bekker & Aldrich (2011) and Bekker (2013) found that the MOO CEM generates high-quality results while requiring relatively few evaluations of the objective functions. Stadler (2012) and Hauman & Bekker (2014) applied the MOO CEM to real-world problems and found that the metaheuristic performs well. Scholtz (2014) studied the metaheuristic by comparing the performance of the MOO CEM as well as two hybrid algorithms based on the MOO CEM to the performance of two algorithms that take relationships between decision variables into account. Importantly, Scholtz (2014) found that on large problems, the MOO CEM as well as a hybrid algorithm based on the MOO CEM typically perform better than algorithms that take the relationships between decision variables into account.

The MOO CEM is a population-based algorithm (as opposed to a single-solution-based algorithm), as such, during each iteration of the algorithm, a number of solutions (as opposed to a single solution) are produced. The MOO CEM produces a population of N solutions by sampling from a unique probability density function for each decision variable in the problem N times. Once a full population of solutions have been generated, the MOO CEM uses a ranking algorithm to evaluate the quality of the solutions that have been generated, based on their objective function values. The best ranked solutions are added to a set of solutions known as "Elite". The probability density function for each decision variable is updated, based on the probability density functions for that variable contained in the Elite set, and a new population of N

solutions (referred to as a new "generation") is generated. In this way, the algorithm increases the likelihood of generating solutions of a superior quality (this is referred to as *exploitation*). In order to ensure that the algorithm does not become trapped in one area of the solution space, the MOO CEM introduces new solutions into the Elite set. This specific mechanism is invoked randomly. The algorithm starts searching in a new area of the solution space (this is referred to as *exploration*).

As discussed in the introduction to this section, both the problem formulation and the metaheuristic were adjusted in order to enable the MOO CEM to be applied to the real-world problem. This process of problem and algorithm adjustment was executed in a number of iterative phases described next. Figure 5.1 provides a comparison of the shape and size of the variable matrices and structures used in each of these iterations.

5.4.1.1 MOO CEM solution phase one

In its original form, the MOO CEM is most commonly applied to continuous rather than discrete problems. The algorithm also does not have the ability to take constraint equations into consideration. During the first MOO CEM solution phase, the focus was on adjusting the problem formulation to eliminate the need for separate constraint equations, and to adjust the functioning of the MOO CEM so that it is able to generate solutions to this problem, evaluate these solutions and exploit good solutions.

The mechanism by which the algorithm ensures *exploration* of the solution space was also adjusted from the standard mechanism that is invoked randomly (described in the preceding subsection). The *exploration* mechanism was adjusted to evaluate the diversity of the solutions contained in the "Elite" set and to generate a new set of probability density functions for each decision variable (based on random number generation) if the diversity fell below a threshold value.

1. The problem was formulated using an allocation matrix with a row for each of the 3 279 DOH facilities and a column for each of the 61 CD4 laboratories as well as a column to denote POC testing (depicted in the top section of Figure 5.1). Constraints were incorporated as desribed next:

   (a) The coverage distance constraint was implemented using the concept of a "big M". Therefore the health impact goal function was penalised by a very large amount whenever a DOH facility's testing was assigned to a CD4 laboratory outside of its coverage distance; and

   (b) Each DOH facility must either be assigned to exactly one laboratory or to POC testing. Therefore each row of the allocation matrix had to contain

exactly one "1". As mentioned, the MOO CEM samples from a unique probability density function to determine the value of each decision variable. Each row of the allocation matrix was set as a decision variable and the MOO CEM was programmed to sample from unique truncated Poisson distributions in order to allocate exactly one "1" to one of 62 positions for each row of the allocation matrix.

2. The metaheuristic now worked as follows:

   (a) The algorithm stored a unique λ value for the truncated Poisson distribution associated with each of the 3 279 decision variables;

   (b) It was possible to generate an infeasible solution by allocating a DOH facility to a laboratory that falls outside its coverage distance. However, as discussed, the concept of a "big M" was employed to ensure that these solutions generated a weak performance in terms of the health impact objective;

   (c) *Exploitation* of good solutions was ensured by adjusting the λ value of each decision variable, based on the λ values in the "Elite" set; and

   (d) As described in the introduction to the subsection, *exploration* of the solution space was implemented by evaluating the diversity of the solutions in the "Elite" set. This was done by converting each of the 3 279 binary strings to a binary representative number and then comparing each representative number for each of the solutions in the "Elite" set.

This model was programmed and implemented in Matlab. The programming and implementation of the mathematical model were verified by running small sets of test data, writing out interim results at various points in the code, duplicating the manipulation of this data in a Microsoft Excel spreadsheet and comparing the results to those generated by the Matlab model. Smaller portions of the Matlab code were also isolated into separate models and tested against an Excel spreadsheet in a similar manner. After the programming of the model had been succesfully verified, it was applied to the full data set and it succesfully generated a Pareto set.

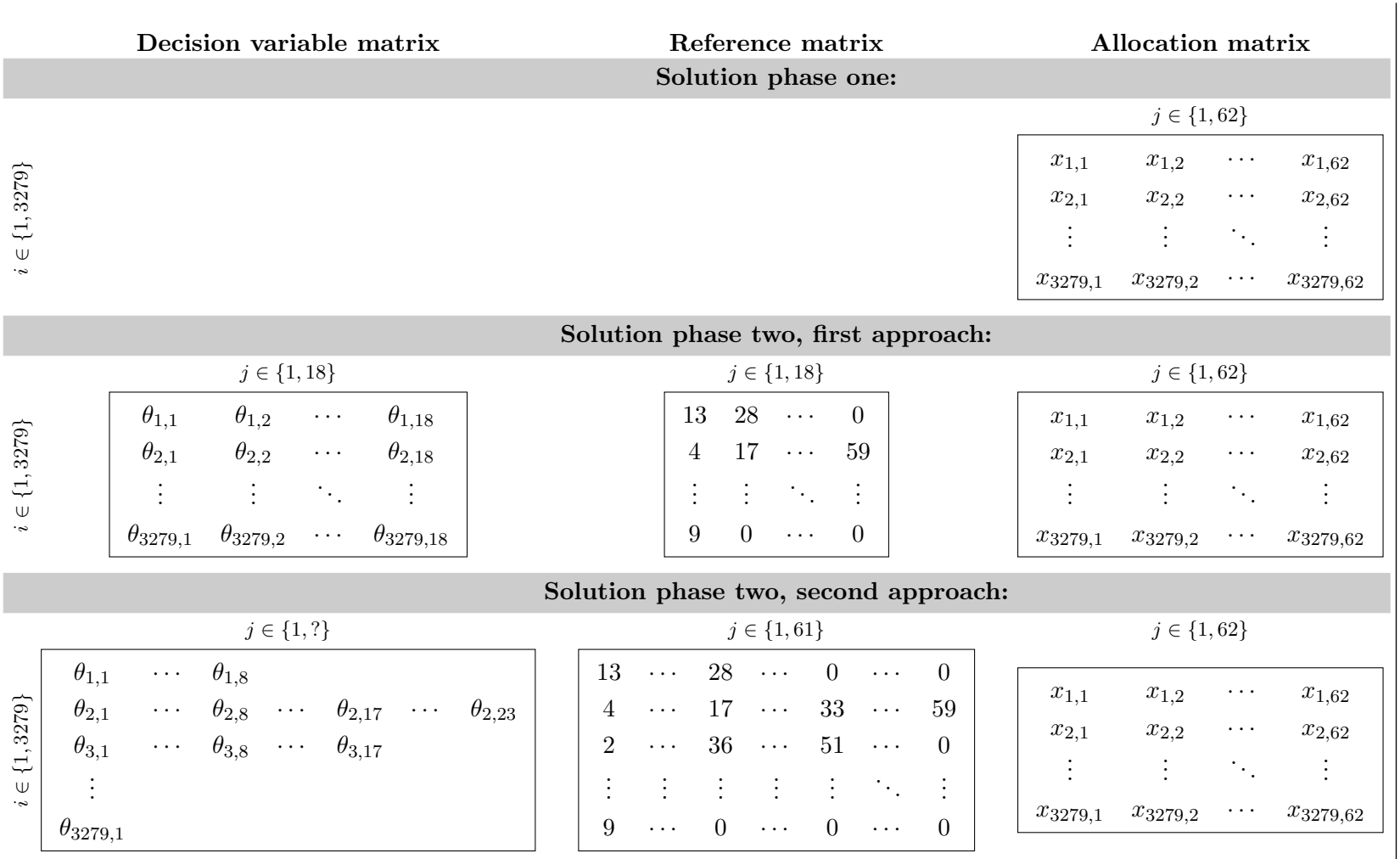

Figure 5.1: Decision variable structure in MOO CEM.

5.4.1.2 MOO CEM solution phase two

In the second MOO CEM solution phase, the focus was on enabling the metaheuristic to solve the real-world problem more efficiently and to generate better quality results. The proposal was that the metaheuristic would be able to (i) find feasible solutions and (ii) *exploit* these feasible solutions more efficiently if the truncated Poisson distribution used to generate the value of each decision variable were shorter. (As described in Subsection 5.4.1.1, each truncated Poisson distribution was defined on the range $\{0, 61\}$ during the first solution phase). The process of incorporating a reference matrix to shorten each truncated Poisson distribution is described next.

1. As a first approach, a standard number of options were made available for each DOH site. In addition to the allocation matrix used in Subsection 5.4.1.1, a decision variable matrix and a reference matrix were also introduced to the problem. The decision variable matrix had significantly fewer columns than the allocation matrix. (Experiments were run with different numbers of standard options for each DOH site, ranging from two options to 18 options.) The reference matrix was used to keep track of which laboratory each cell in the decision variable matrix referred to and this was used to convert the decision variable matrix to the allocation matrix before evaluating the goal functions. (Refer to the middle section of Figure 5.1.) This approach was verified using a similar method to that described in Subsection 5.4.1.1 and a Pareto set was succesfully generated for the full data set. This approach was not ideal because:

   (a) for DOH facilities with a large number of CD4 laboratories within their coverage distance, all of the feasible options were not being considered; and

   (b) for DOH facilities with only a small number of laboratories in their coverage distance, infeasible options were also being considered (the "big M" approach described in Subsection 5.4.1.1 was used to penalise the health impact objective in these cases).

2. As a second approach, the algorithm was adjusted so that the exact number of feasible options that exist for each DOH site[1] was made available to that site. This was achieved by adjusting the decision variable structure (introduced in the preceding paragraph) so that it had varying row lengths. (This is depicted in the bottom section of Figure 5.1.) For each of the 3 279 decision variables, the

[1] The number of CD4 laboratories that fall within the coverage distance of each DOH site as well as the POC option.

metaheuristic now kept track of both the length and the λ-value of the truncated Poisson distribution. While the λ-value was adjusted as the metaheuristic explored good solutions, the length of each truncated Poisson distribution remained fixed. The implementation of this approach was verified using a similar method to that described in Subsection 5.4.1.1 and a Pareto set was generated for the full data set. This approach was superior to those implemented previously because:

(a) all feasible solutions are considered;

(b) no infeasible solutions are considered; and

(c) due to the reduced size of the decision variable structure, the metaheuristic is able to generate better quality solutions with less computational effort.

5.4.1.3 MOO CEM solution phase three

During the third MOO CEM solution phase, the focus was on ensuring that the metaheuristic thoroughly *explored* the solution space. Specifically, different methods for generating the initial set of λ values during each main iteration of the metaheuristic were experimented with. The details of this experimentation will not be covered here but an important conclusion from this experimentation is that the functioning of the MOO CEM can be greatly affected by the method used to generate the probability density function parameters. As an illustration of this point, earlier experimentation with the generation of probability density function parameters was succesful in directing the metaheuristic to explore a specific portion of the solution space, but could not succeed in widening the scope of this exploration. Consequently, an approach was formulated where the metaheuristic was run 22 times, directed to explore a specific portion of the solution space during each run. (Due to the stochastic nature of the metaheuristic, it was challenging to direct the search at a specific portion of the solution space. The 22 runs were decided on through experimentation, and were selected with the aim of ensuring that the metaheuristic explores the entire solution space.) These results were then combined to form a final Pareto set. This was the status quo for approximately two months until further experimentation with alternative methods of generating probability density function parameters eventually led to a method that allowed the metaheuristic to explore the solution space as completely during only one run as it had previously done over the course of 22 runs.

5.4.1.4 Other factors considered during the MOO CEM solution

Before concluding the discussion of the MOO CEM, it is important to mention two parameters that influence the working of the metaheuristic:

1. The ϵ parameter governs the threshold value for an acceptable level of diversity in the "Elite" set. As discussed in Subsection 5.4.1, this ensures thorough *exploration* of the solution space.

2. The α parameter governs the extent to which the probability density function for each decision variable is updated based on the solutions contained in the "Elite" set. As discussed in Subsection 5.4.1, this ensures thorough *exploitation* of good solutions.

During all three phases of the MOO CEM solution, experiments were conducted with different combinations of these parameter values in order to determine combinations that generated good results in a short period of time. Parameter adjustment of this type is not an exact science and, though this experimentation was documented and conducted in a systematic manner, it was not the purpose of this research to conduct exhaustive experiments on these parameters of the MOO CEM. In addition to the ϵ and α parameters, other settings of the metaheuristic such as the population size and the maximum number of evaluations that are executed were also experimented with.

5.4.1.5 Results generated by the MOO CEM

Due to the need for sensitivity analysis based on the coverage distance (discussed in Subsection 5.1.9), three scenarios were solved using the MOO CEM metaheuristic. The results are displayed in Figure 5.2. As shown, the real-world problem does not seem to display a significant sensitivity to the different coverage distance values (100km, 150km and 200km) that are considered. It is therefore recommended that it is reasonable to proceed with further analysis based on a coverage distance of 150km.

5.4.2 Single-objective solution with a commercial algorithm

As a method to verify the quality of the solutions generated by the MOO CEM, a decision was made to apply a standardised Matlab solver or solvers to a single-objective version of the problem. (The single-objective version of the problem described in the introductory paragraphs of Section 5.4 was used, therefore the maximum solution cost was constrained and the health impact was set as the objective.) The complete set of

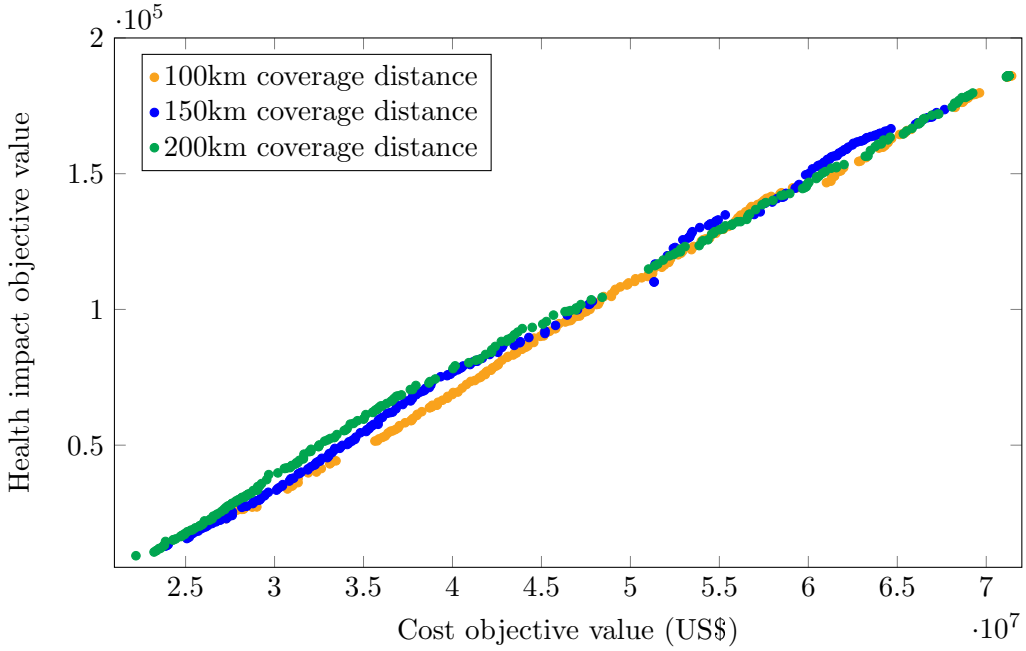

Figure 5.2: The MOO CEM results.

solvers available in both the Matlab Optimisation Tool and the Matlab Global Optimisation Tool are summarised in Table 5.5. As described in the table, four of the solvers can be applied to a single-objective version of the real-world problem.

The standardised Matlab solvers require that the problem constraints be formulated using two equations

$$A.x \leq b, \tag{5.9}$$

$$Aeq.x = beq. \tag{5.10}$$

Thus all linear *inequality* constraint equations are grouped into a single matrix (A and b respectively) for the left and the right hand side of (5.9) and all linear *equality* constraint equations are similarly grouped into a single matrix (Aeq and beq) for the left and the right hand side of (5.10). The objective function as well as any nonlinear constraint(s) that may exist are defined as functions that are programmed in the normal Matlab script environment. The solvers also generally allow the user to set a lower and an upper bound for the variable(s) and to define an initial value for the variable(s). The first step in solving a single-objective version of the real-world problem using one of the standardised Matlab solvers, was to convert the mathematical formulation given in Subsection 5.2.5 to suit the standardised problem format required by the Matlab

Table 5.5: Suitability of Matlab solvers.

Matlab solver name	**Suitability for application to real-world problem**
Binary integer programming ("bintprog")	Unsuitable: Unable to handle nonlinear constraints
Multi-objective goal attainment ("fgoalattain")	Unsuitable: Multi-objective optimisation
Single-variable nonlinear minimisation with bounds ("fminbind")	Unsuitable: Unable to handle multiple variables
Constrained nonlinear minimisation ("fmincon")	Suitable
Minimax optimisation ("fminimax")	Suitable
Unconstrained nonlinear minimisation ("fminsearch")	Unsuitable: Unable to handle constraints
Unconstrained nonlinear minimisation (fminunc")	Unsuitable: Unable to handle constraints
Semi-inifite minimisation ("fseminf")	Unsuitable: the problem does not have semi-inifinite constraints
Nonlinear equation solving ("fsolve")	Unsuitable: Unable to handle constraints
Single-variable nonlinear equation solving ("fzero")	Unsuitable: Unable to handle multiple variables
Genetic Algorithm ("ga")	Suitable
Multi-objective optimisation using Genetic Algorithm ("gamultiobj")	Unsuitable: Unable to handle nonlinear constraints
Linear programming ("linprog")	Unsuitable: Unable to handle nonlinear constraints
Nonlinear curve fitting ("lsqcurvefit")	Unsuitable: Curve fitting is not applicable to the problem
Constrained linear least squares ("lsqlin")	Unsuitable: Unable to handle nonlinear constraints
Nonlinear least squares ("lsqnonlin")	Unsuitable: Unable to handle constraints
Nonnegative linear least squares ("lsqnonneg")	Unsuitable: Unable to handle constraints
Pattern Search ("patternsearch")	Suitable
Quadratic programming ("quadprog")	Unsuitable: Unable to handle nonlinear constraints
Simulated annealing algorithm ("simulannealbnd")	Unsuitable: Unable to handle constraints

solvers.

As a first attempt to convert the real-world problem to a suitable format, an allocation matrix with 3 279 rows and 62 columns was used as the decision variable (x) and the constraint matrices (A, b, Aeq, and beq) were set up based on this allocation matrix. The accuracy of this model formulation was tested by applying one of the suitable solvers to smaller versions of the real-world problem. The results that were generated were written out to Excel files and were analysed to verify that all constraint equations were functioning correctly. Following this verification step, an attempt was made to solve the full problem. Unfortunately, Matlab was not able to solve the real-world problem in this formulation as the size of the constraint matrices that were generated were larger than Matlab's memory allows for.

As a second attempt to convert the real-world problem to a suitable format, a decision variable vector with 3 279 rows and 1 column was used in combination with a reference matrix. (This is similar to the approach combining a decision variable and a reference matrix used in Subsection 5.4.1.2, although the decision variable used in Subsection 5.4.1.2 was a structure with varying numbers of columns rather than a vector.) The constraint equations were adjusted in accordance with the decision variable format, in this format, it was possible to formulate the mathematical model using only a nonlinear constraint function. Therefore no matrices for the linear equality and inequalities were required. The accuracy of this model formulation was tested using small data sets and analysis in Excel, similar to the approach used previously. Following this verification step, an attempt was made to solve the full problem with each of the four suitable solvers identified in Table 5.5. In each case, an initial feasible solution was provided as a starting point for the solver:

1. The "fmincon" constrained nonlinear minimisation solver ran for approximately 40 minutes without generating any feasible solutions. The solver then generated an error message. Therefore, though this solver is suited to all of the characteristics of the single-objective version of the real-world problem, it was not able to solve the problem efficiently in the formulation described here;

2. The "fminimax" minimax optimisation solver responded in exactly the same manner as the "fmincon" solver;

3. During the first hour, the Genetic Algorithm solver generated several feasible results that had a better objective function value than the initial feasible solution. This confirmed that the solver was suited to all of the characteristics of the single-objective version of the real-world problem and that it had the potential to solve

the problem formulation described here efficiently; and

4. The Pattern Search solver ran for an hour without generating any feasible solutions or any error messages. Therefore it is reasonable to conclude that, though this solver is compatible with all of the characteristics of the real-world problem, it is not able to solve the problem efficiently in the formulation described here.

From the experimentation above, a decision was made to proceed using only the Genetic Algorithm solver. The Matlab Genetic Algorithm has various parameters that can be adjusted in order to improve the performance of the solver. The following parameters were experimented with: population size, reproduction crossover fraction, migration fraction, migration interval, stopping criteria time limit, and stopping criteria stall function value tolerance. Similar to the approach used during the MOO CEM parameter experimentation, though this experimentation was documented and conducted in a systematic manner, the aim was not to perform exhaustive experiments on the effects of different combinations of parameter values. The purpose of solving a single-objective version of the real-world problem using a commercial algorithm, is to verify the quality of the results generated by the MOO CEM. The verification process is described in more detail in the next subsection.

5.5 Verification of the solution methodology

Two approaches were followed to verify the quality of the solution generated by the MOO CEM: (i) the cost of the status quo (where no POC testing is used) as well as the cost and incremental health impact of a scenario where POC testing is used as widely as possible were estimated using deterministic methods; and (ii) a single-objective version of the problem was solved using a commercial algorithm. A coverage distance of 150km was used during verification.

5.5.1 First verification step

The status quo is equivalent to the scenario with the lowest possible cost and has an *incremental* health impact of zero. (As stated in Subsection 5.2.1, facilities that make use of laboratory-based testing are assumed to have no incremental health impact.) The scenario where POC testing is used as widely as possible, is the scenario with the highest possible cost and incremental health impact. The objective function values for these scenarios were estimated to establish whether the MOO CEM explored the

full range of options for the incorporation of POC testing into the diagnostic service delivery model.

The costing of the status quo scenario is based on the most conservative estimate of the current annual cost of delivering CD4 testing to the South African public healthcare sector. Two approaches were considered, both approaches used the number of CD4 tests performed by the NHLS in the 2012/2013 financial year (refer to Subsection 5.1.3) as part of the calculation:

1. Use the average cost per test for the laboratory tiers (tiers 3, 4 and 5) as given in Table 5.2. This results in a cost per test of US$6.34, which translates to a total annual testing cost of US$21 291 280; or

2. Use the average number of tests per facility to determine which laboratory tier's cost per test amount to use in the calculation. The average number of tests per facility per day is 229, this falls within the tier 4 laboratory classification (as given in Table 5.2). A total annual testing cost of US$20 955 455 was estimated, based on the tier 4 cost per test of US$6.24.

The second approach was selected as it generated the lowest estimate of the status quo cost. (As this cost will be used to determine the lowest cost in the solution space, this is the most conservative approach.)

The estimates for the scenario where POC testing is used as widely as possible were developed by asigning POC testing to all ART initiation facilities where the total annual testing volume is less than the threshold for POC testing defined in Table 5.2. The remaining 20 ART initiation facilities were allocated to one of three CD4 laboratories. This approach generated an estimated testing cost of US$71 215 960 and an incremental health impact of 186 013 people.

These two scenarios are compared to the results generated by the MOO CEM in Figure 5.3. Though it is clear that the MOO CEM does not explore the most extreme reaches of the solution space, it does appear to explore a sufficiently wide range of options from very limited use of POC (at a cost of US$23 825 197) to widespread use of POC (at a cost of US$67 656 318).

5.5.2 Second verification step

In the second verification step, the Matlab Genetic Algorithm was used to solve a single-objective version of the real-world problem. The incremental health impact was set as the only objective and the cost function was converted to an inequality constraint.

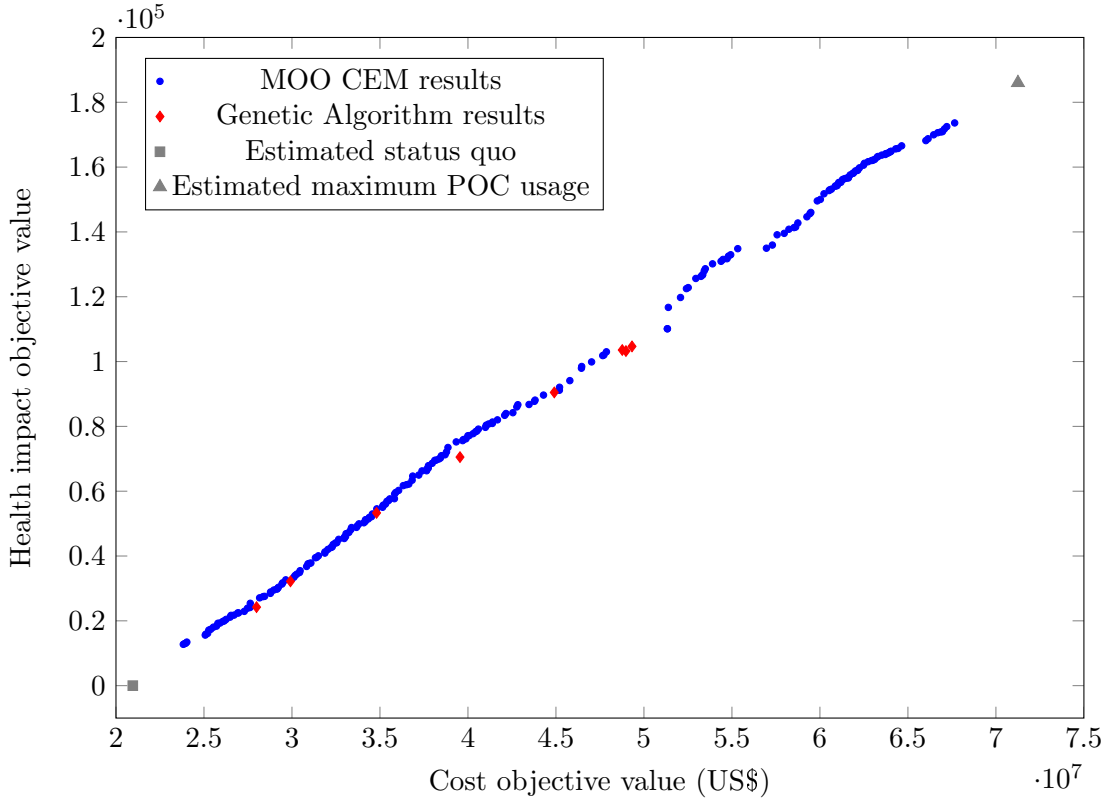

Figure 5.3: Comparison of MOO CEM and Matlab Genetic Algorithm solution quality.

In order to generate several points that could be compared with the Pareto set, the optimisation was run 10 times with the maximum allowable testing cost set to a different value each time. (The maximum cost was varied from US$25 million to US$70 million at intervals of US$5 million.) The Matlab solver was set to terminate either after eight hours or if all of the solutions in 15 consecutive generations had failed to improve on the objective function value.

The points generated by the Matlab Genetic Algorithm solver are compared to the Pareto set generated by the MOO CEM in Figure 5.3. From the figure, it is evident that the Matlab Genetic Algorithm struggles to explore both the low cost, low health impact and the high cost, high health impact portions of the solution space. However, as shown, the quality of the results generated by the MOO CEM compare well with that of the results generated by the commercial algorithm. The figure also clearly shows that the MOO CEM struggles to explore the portion of the solution space with a cost of approximately US$50 million. However, the Matlab Genetic Algorithm results show that solutions do exist in this portion of the solution space. An important conclusion that can be made from this is that one could generate a more complete set of solutions to the real-world problem by combining solutions that have been generated by both the MOO CEM and the commercial algorithm into a single data set and then generating a Pareto set from this combined data set.

5.6 Final solution set

With reference to the conclusion made in the previous subsection, the solutions presented here have been drawn from a Pareto set that was formed based on both the MOO CEM and the commercial algorithm's results. As stated in Section 1.2, the aim of this research is to provide decision-makers with information on a range of scenarios for the use of POC CD4 testing in South Africa. To accomplish this, the details on a set of nine solutions ranging from a budget of US$25 million to US$65 million at intervals of US$5 million are given. (US dollars was selected as the currency for the modelling as it is the currency that was used in the Cassim *et al.* (2014) study on which the costing in the model is based.)

Figure 5.4 contains both the MOO CEM Pareto set solutions with a system cost that is closest to the defined budget as well as the Matlab Genetic Algorithm results for each of the defined budgets. The budgets were not intended as absolute constraints, therefore the Matlab Genetic Algorithm was set up to generate solutions with a total system cost that did not exceed the budget amount by more than a tolerance amount. Similarly,

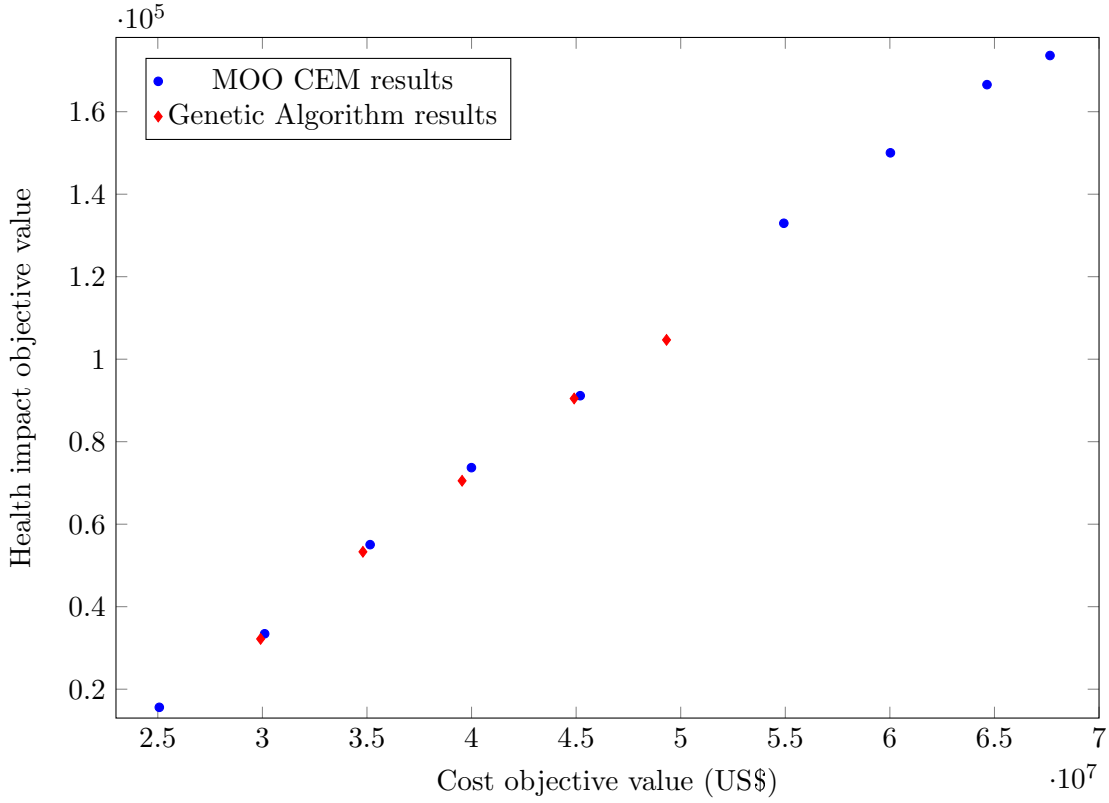

Figure 5.4: Final MOO CEM and Matlab Genetic Algorithm solution sets.

the MOO CEM solution with a total testing cost that was *closest* to the budget amount was selected, provided the testing cost did not exceed the budget amount by more than US$1 million. The MOO CEM Pareto set did not contain a solution with a total testing cost of between US$49 million and US$51 million.

Table 5.6 gives the summarised final solution set that will be discussed in more detail in Chapter 6. For a budget of US$50 million, the Matlab Genetic Algorithm solution was selected. The MOO CEM solutions were selected for the remainder of the budget options being explored.

Table 5.6: Final solution set.

Scenario	Budget	Total system cost	Incremental health impact	Solution method
1	$25 million	$25 071 516	15 603 people	MOO CEM
2	$30 million	$30 108 463	33 433 people	MOO CEM
3	$35 million	$35 150 572	55 040 people	MOO CEM
4	$40 million	$39 993 712	73 708 people	MOO CEM
5	$45 million	$45 195 168	91 149 people	MOO CEM
6	$50 million	$49 322 233	104 678 people	Matlab GA
7	$55 million	$54 930 000	132 954 people	MOO CEM
8	$60 million	$60 025 574	150 032 people	MOO CEM
9	$65 million	$64 643 600	166 558 people	MOO CEM

5.7 Conclusion: Modelling the real-world problem

This chapter discussed the modelling of the real-world problem by describing the input data analysis, presenting the mathematical formulation and describing the solution methodology. The next chapter will discuss the results that have been generated.

Chapter 6

Analysis of results

The previous chapter detailed the modelling of the real-world problem. This chapter provides an analysis of the results obtained for the real-world problem and discusses the implication of these results in terms of the healthcare delivery system.

6.1 Presentation of results

Arguably the most important contribution of this research is to provide decision-makers within the SA public healthcare sector with an indication of where the optimal trade-off points between an improved healthcare outcome and increased healthcare spending lie. This information better equips individuals involved in managing diagnostic service delivery and those responsible for managing the allocation of public funds by providing an indication of the alternative scenarios that should be considered.

The final solution set was presented in Table 5.6. When comparing these scenarios, there are a number of factors other than the total testing cost and the incremental health impact of the solution that decision-makers should take into account:

1. The number of ART initiation sites where POC is implemented;
2. The number of CD4 laboratories that are required;
3. The percentage of the total testing workload that is executed on POC testing devices; and
4. The incremental cost per additional person initiated onto treatment.

Table 6.1 gives the complete data set for each of the nine selected scenarios as well as the status quo (scenario 0) and a scenario where POC testing is implemented as widely as possible (scenario 10). The calculation of the cost and the incremental health impact for these two scenarios was described in Subsection 5.5.1.

Table 6.1: Summarised results to the real-world problem.

Scenario	Annual CD4 testing cost	Incremental health impact	Number of POC sites	Number of CD4 laboratories	% of testing volume performed using POC	Incremental cost per additional person initiated
0	$20 955 455	0	0	61	0%	$0
1	$25 071 516	15 603	453	61	7%	$264
2	$30 108 463	33 433	739	61	16%	$274
3	$35 150 572	55 040	1 215	57	25%	$257
4	$39 993 712	73 708	1 480	54	34%	$258
5	$45 195 168	91 149	1 757	54	43%	$266
6	$49 322 233	104 678	1 743	33	55%	$271
7	$54 930 000	132 954	2 236	47	64%	$256
8	$60 025 574	150 032	2 618	44	71%	$260
9	$64 643 600	166 558	2 827	35	81%	$262
10	$71 251 960	186 013	3 259	4	91%	$270

From an operational perspective, the number of ART initiation sites where POC testing will be utilised is an important metric as it would be reasonable to conclude that the operational management of the healthcare system would become more complex as testing is executed at a larger number of facilities. Similarly, the number of CD4 laboratories that are required for each scenario is an important operational consideration. It is logical that, as the use of POC increases, a reduced number of CD4 laboratories will be required to deliver diagnostic services. This would simplify the operational management of the laboratory network by reducing the number of laboratory sites that require staff with specialist skills to perform CD4 testing, reducing the number of laboratory sites where specialist CD4 testing equipment needs to be maintained, etc.

The interpretation of the incremental cost per person initiated is discussed in more detail in Section 6.2. This value was calculated by subtracting the total testing cost of the status quo from the total testing cost of each scenario. This amount was then divided by the total number of additional people initiated onto ART in each scenario (n_i in (5.1)) in order to determine the incremental CD4 testing cost associated with initiating each of these additional people.

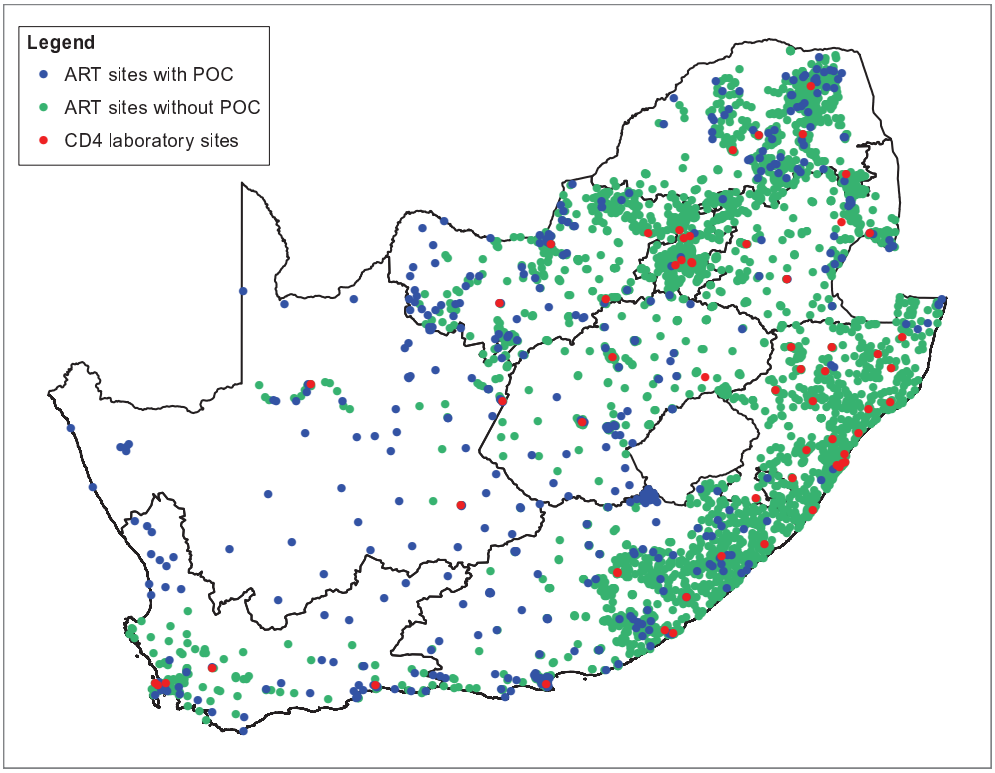

Figure 6.1: Scenario 1 – Location of POC ART initiation sites, non-POC ART initiation sites and NHLS CD4 laboratories.

Figures 6.1 to 6.3 visually depict the distribution of POC sites for scenarios 1, 5 and 9. One can clearly see how the number of POC sites increase from scenario 1 to scenario 9.

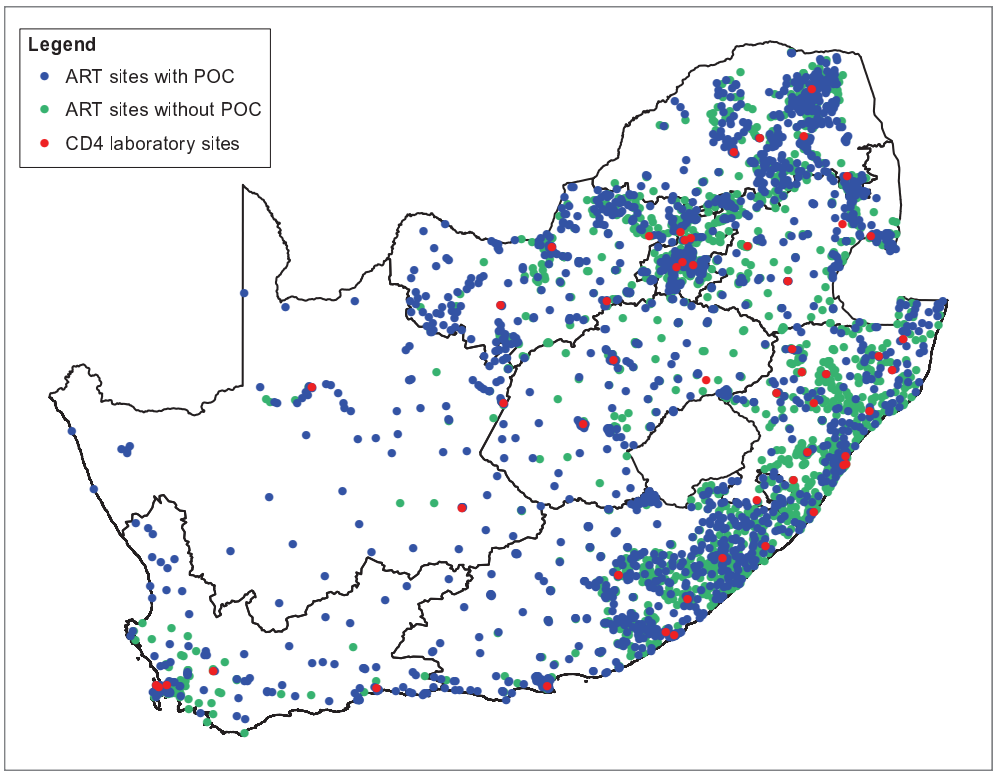

Figure 6.2: Scenario 5 – Location of POC ART initiation sites, non-POC ART initiation sites and NHLS CD4 laboratories.

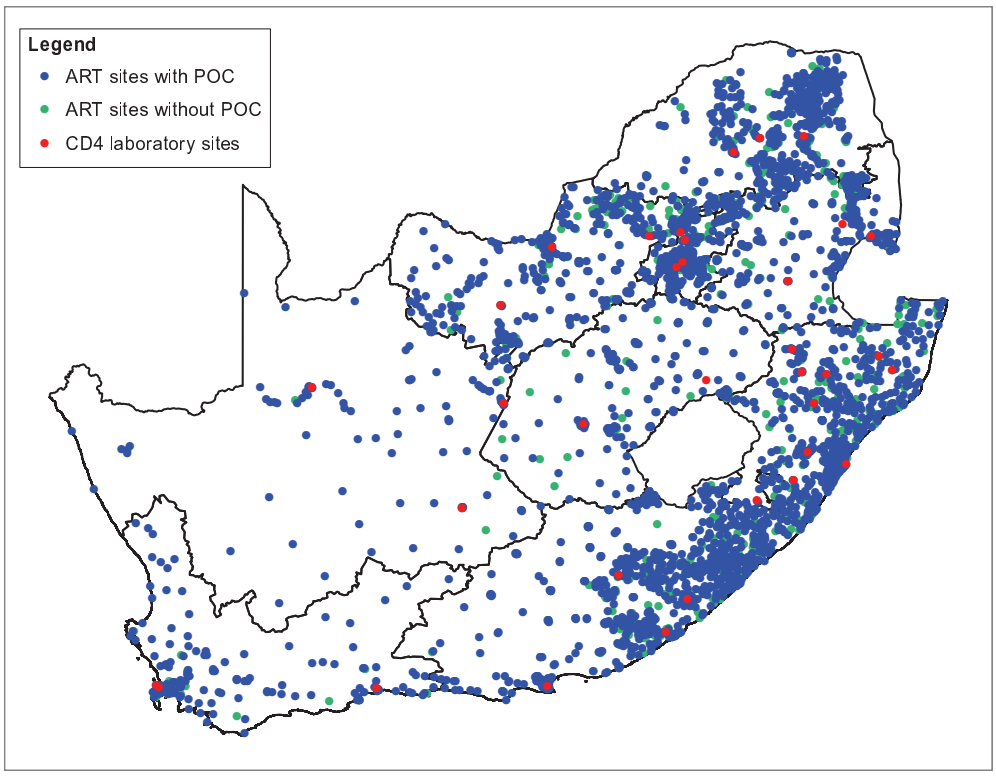

Figure 6.3: Scenario 9 – Location of POC ART initiation sites, non-POC ART initiation sites and NHLS CD4 laboratories.

6.2 Analysis of results

It would be reasonable to conclude that the available healthcare budget in the country would play a large role when decision-makers evaluate these alternative scenarios. However, in addition to the available budget, it would be important to also consider (i) the *cost effectiveness* of each scenario; as well as (ii) the *total public healthcare expenditure on HIV*.

6.2.1 Cost effectiveness of the scenarios

The incremental cost per additional person initiated is an indication of the cost effectiveness of each scenario. As shown in Table 5.6 the low cost scenarios have a lower incremental cost per additional person initiated than the high cost scenarios. To provide some perspective on these costs: based on the most conservative estimate of the current laboratory-based CD4 testing cost (scenario 0), the status quo CD4 testing cost per person initiated is US$33.72. Acoording to DHIS data[1], 621 379 individuals were initiated onto ART at this cost during the 2012/2013 financial year.

6.2.2 Total healthcare system cost

Though estimates of the exact amounts vary from one study to another, literature agrees that the total annual cost to the healthcare system of a patient that is eligible for ART and that is in fact receiving the treatment is lower than that of a patient that is eligible but is not receiving the treatment. Therefore, though the expenditure on CD4 testing in scenarios 1 to 10 would be higher than at present, it is possible that the total public healthcare expenditure on HIV would be lower. In order to make a recommendation on the effect of each of these scenarios on the total health system cost, one would need to (i) systematically review the literature to establish which of the published healthcare system cost estimates to use; and (ii) use modelling to calculate the nett year-on-year financial effect of an increased number of patients being initiated onto ART each year. Such a study falls outside the scope of this thesis, but it would be valuable as the topic of further research.

6.3 Other considerations

In addition to the quantified financial and healthcare impact characteristics of each scenario, there are also a number of more practical considerations which one would

[1] Data set introduced in Subsection 5.1.8.

need to consider before implementing any of the scenarios presented here. Some of these practical considerations are listed here in order to give an indication of the types of factors that should be considered, however, this is by no means a comprehensive list.

1. Is the necessary infrastructure (most importantly a reliable electricity supply) available at the ART initiation facilities that have been selected for the implementation of POC testing;

2. A decision on who would take the responsibility for the execution of POC testing: would DoH staff be responsible for executing this work, or would NHLS staff be allocated to the ART initiation facilities to execute the work;

3. Other effects that POC testing would have on the workload of clinic staff – for example, patients will be coming to the clinic at least one less time before being placed on treatment.;

4. The need to manage the quality of the results generated by POC instruments. For example, one would need to ensure that analyser calibrations take place as scheduled and that the necessary corrective actions are taken when calibration fails. It is reasonable to assume that neither the NHLS nor the DoH would be willing to take responsibility for signing off the test results if such a system were not in place as this could have medico-legal implications; and

5. The effect of POC testing on the accuracy of HIV surveilance data in South Africa. At present, these analysers are not set up to automatically submit the result of each test to a centralised database. This is something that the laboratory-based analysers are set up to do.

It is also recommended that decision-makers consider revisiting one of the business decisions that have informed this study. It is likely that revisiting the decision *not to relocate any of the existing CD4 laboratories* may enable one to deliver HIV services to public healthcare users in the country significantly more cost-effectively. As highlighted in Figure 2.3, NHLS laboratory sites are not distributed uniformly accross the country. The maps indicating the location of CD4 laboratories in Figures 6.1 to 6.3 also illustrate this unbalanced distribution. It is recommended that it would be prudent to consider relocating some of the existing laboratory-based CD4 testing capability so that a larger number of ART initiation sites fall within the coverage distance of a laboratory.

Lastly, it is important that decision-makers understand the limits of the modelling work that has been done here. As discussed in Section 5.4, this is a difficult problem

to solve. During this study, the problem was solved using a metaheuristic that has been designed for multi-objective optimisation as well as a commercial algorithm. It is recommended that, for a problem of this size and nature, one might manage to generate superior quality results by applying a customised heuristic that has been specifically designed for application to this real-world problem.

6.4 Conclusion: Analysis of results

This chapter presented an analysis of the results obtained from the modelling of the real-world problem. The next chapter will provide a summary of the project and serve as a conclusion to the project report.

Chapter 7

Summary and conclusions

The previous chapter discussed the results obtained from the modelling and solving of the real-world problem. This chapter provides a summary of the research that has been done before presenting the research findings and making suggestions for further research.

7.1 Project summary

The purpose of this research was to apply Operations Research to solve a location science problem concerned with the provision of HIV diagnostic services in the South African public healthcare sector. This was done by modifying a standardised mathematical formulation of a location science problem in order to ensure that it was able to take all of the characteristics of the real-problem into consideration. Both a metaheuristic for multi-objective optimisation and a commercial algorithm for single-objective optimisation were then applied to solve the real-world problem.

As background to the research, Chapter 2 provided a brief overview of the status quo of the South African public healthcare sector, the strategic aims and objectives of the sector, and the current diagnostic service delivery model. The real-world problem was introduced by describing the role of CD4 testing in the ART initiation pathway and presenting research on loss to follow-up and the expected impact of POC CD4 testing on this phenomenon.

The discipline of Operations Research as well as the field of location science were introduced in Chapter 3. The need for the application of Operations Research to support decision-making in healthcare was established. The lack of literature on the application of location science to real-world problems was confirmed and the applicability of location science to the real-world problem was demonstrated.

Eight standard types of location science models were identified and their mathematical formulations presented in Chapter 4. These standard models were analysed against the characteristics of the real-world problem and the most compatible model was selected.

The modelling of the real-world problem was described in Chapter 5. The available input data was analysed and processed for use in the model; the mathematical formulation of the standard location science model selected in the previous chapter was adjusted according to the available input data and other characteristics of the real-world problem; and the problem was solved using both the MOO CEM metaheuristic and a commercial Matlab algorithm.

Chapter 6 presented the results obtained from the modelling of the real-world problem. The results were analysed in terms of their cost-effectiveness. The importance of evaluating the impact of a decision on the use of POC CD4 testing on the total healthcare system cost was highlighted. Examples of practical considerations that decision makers would need to take into account when making a decision on the use of POC CD4 testing were given. Finally, business decisions that informed this modelling, but that do not necessarily lead to the most effective solution for diagnostic service provision in SA were outlined.

7.2 Research findings

The research produced the following three findings:

1. It highlighted the importance of collecting accurate healthcare data in order to allow for informed decision-making on healthcare interventions (specifically in the public healthcare sector);

2. It demonstrated the challenges associated with applying theoretical model formulations to complex, real-world problems; and

3. It demonstrated that, for multi-objective problems of this size and level of complexity, combining the results from the multi-objective problem with those of a single-objective version of the problem, may generate a superior set of final results.

7.3 The contributions of this research

This research contributes to both academic literature and to South African public healthcare management:

1. It adds to the location science application literature by applying Operations Research to a novel real-world testing device location problem;

2. It adds to the location science literature by proposing a framework for selecting the most suitable standard mathematical model for application to a real-world problem;

3. It adds to the metaheuristic literature by describing a tailored method for applying the MOO CEM to an allocation problem;

4. It provides healthcare decision makers in South Africa with quantified data on the optimal trade-off between cost and health impact of the various scenarios for the implementation of POC CD4 testing that should be considered; and

5. It provides detailed solutions for CD4 testing provision at 3 279 ART initiation facilities as well as the resulting optimal referral network to 61 CD4 laboratories for each of the nine proposed scenarios.

7.4 Opportunities for further work

It is proposed that the following opportunities for further work be explored. It is believed that each of these suggested studies would add value to the quality of decision-making in the SA public healthcare sector:

1. Revisit the business decision on the relocation of CD4 testing capacity to NHLS laboratories that do not currently offer it. Generate scenarios that incorporate a simultaneous use of POC testing and optimal re-allocation of CD4 laboratory sites;

2. Evaluate the long-term holistic financial impact of each scenario on the healthcare system. Specifically, take the annual treatment cost of HIV positive individuals that are eligible for ART and are either receiving the treatment or not into consideration. Also attempt to translate this holistic financial impact to an annual healthcare provision cost per quality adjusted life year[1] (QALY) in order to allow the options to be evaluated according to a recognised methodology; and

[1] A measure that is typically used in economic evaluations of different healthcare interventions. The cost per QALY is compared to the GDP per person to determine how cost-effective a healthcare intervention is.

3. Attempt to develop a customised heuristic for solving this real-world problem and evaluate whether this heuristic is able to generate better quality results than the MOO CEM and the Matlab GA.

References

Abdool Karim, S.S., Churchyard, G.J., Karim, Q.A. & Lawn, S.D. (2009). HIV infection and tuberculosis in South Africa: an urgent need to escalate the public health response. *Lancet*, **374**, 921–33. 12

Alsalloum, O.I. & Rand, G.K. (2006). Extensions to emergency vehicle location models. *Computers & Operations Research*, **33**, 2725–2743. 57

Anderson, D.A., Crowe, S.M. & Garcia, M. (2011). Point-of-care testing. *Current HIV/AIDS Reports*, **8**, 31–7. 22, 25, 30, 32

Anthonissen, C.A. (2007). *A Population-based Approach to Sequential Ordering Problems*. Master's thesis, Stellenbosch University. 42

Araz, C., Selim, H. & Ozkarahan, I. (2007). A fuzzy multi-objective covering-based vehicle location model for emergency services. *Computers & Operations Research*, **34**, 705–726. 55

Bekker, J. (2012). *Applying the Cross-Entropy Method in Multi-Objective Optimisation of Dynamic Stochastic Systems*. Ph.D. thesis, Stellenbosch University. 107

Bekker, J. (2013). Multi-objective buffer space allocation with the cross-entropy method. *International Journal of Simulation Modelling*, **12**, 50–61. 108

Bekker, J. & Aldrich, C. (2011). The cross-entropy method in multi-objective optimisation: An assessment. *European Journal of Operational Research*, **211**, 112–121. 108

Beraldi, P. & Bruni, M. (2009). A probabilistic model applied to emergency service vehicle location. *European Journal of Operational Research*, **196**, 323–331. 57

Beraldi, P., Bruni, M. & Conforti, D. (2004). Designing robust emergency medical service via stochastic programming. *European Journal of Operational Research*, **158**, 183–193. 56

Boloori Arabani, A. & Farahani, R.Z. (2012). Facility location dynamics: An overview of classifications and applications. *Computers & Industrial Engineering*, **62**, 408–420. 51, 53

Boyle, D.S., Hawkins, K.R., Steele, M.S., Singhal, M. & Cheng, X. (2012). Emerging technologies for point-of-care CD4 T-lymphocyte counting. *Trends in Biotechnology*, **30**, 45–54. 1, 2, 30, 32, 33

Brandeau, M.L., Sainfort, F. & Pierskalle, W.P. (2004). Health care delivery: Current problems and future challenges. In M.L. Brandeau, F. Sainfort & W.P. Pierskalla, eds., *Operations Research and Healthcare: A Handbook of Methods and Applications*, chap. 1, 1–14, Kluwer Academic Publishers, Boston. 44

Canbolat, M.S. & von Massow, M. (2011). Locating emergency facilities with random demand for risk minimization. *Expert Systems with Applications*, **38**, 10099–10106. 57

Cassim, N., Coetzee, L., Schnippel, K. & Glencross, D. (2014). Estimating costs of an integrated tiered CD4 service including laboratory and point of care testing in a remote health district in South Africa. *PLoS ONE*, **Submitted**. 96, 97, 102, 103, 106, 121

Cho, C.J. (1998). An Equity-efficiency Trade-off Model for the Optimum Location of Medical Care Facilities. *Socio-Economic Planning Sciences*, **32**, 99–112. 55

Chopra, M., Lawn, J.E., Sanders, D., Barron, P., Abdool Karim, S.S., Bradshaw, D., Jewkes, R., Abdool Karim, Q., Flisher, A.J., Mayosi, B.M., Tollman, S.M., Churchyard, G.J. & Coovadia, H. (2009). Achieving the health Millennium Development Goals for South Africa: challenges and priorities. *Lancet*, **374**, 1023–31. 9, 10, 12

Church, R.L. (1999). Location modelling and GIS. In P. Longley, M. Goodchild, D. Maguire & D. Rhind, eds., *Geographical information systems: Principles, techniques, applications and management*, 293–303, John Wiley & Sons, New York, 2nd edn. 60, 61

Cocking, C., Flessa, S. & Reinelt, G. (2012). Improving access to health facilities in Nouna district, Burkina Faso. *Socio-Economic Planning Sciences*, **46**, 164–172. 56

Complex Systems Modelling Group (2010). *Modelling in healthcare*. American Mathematical Society, Providence. 43

Cook, S.A. (1971). The complexity of theorem-proving procedures. *Proceedings of the third annual ACM symposium on Theory of Computing*, 151–158. 42

Current, J., Min, H. & Schilling, D. (1990). Multiobjective analysis of facility location decisions. *European Journal of Operational Research*, **49**, 295–307. 81

Current, J., Daskin, M. & Schilling, D. (2002). Discrete network location models. In Z. Drezner & H. Hamacher, eds., *Facility location: Applications and theory*, chap. 3, 81–118, Springer Verlag, Berlin. 48, 49, 50, 51, 53, 61, 71, 72, 76

Daskin, M.S. & Dean, L.K. (2004). Location of health care facilities. In M.L. Brandeau, F. Sainfort & W.P. Pierskalla, eds., *Operations Research and healthcare: A handbook of methods and applications*, chap. 3, 43–76, Kluwer Academic Publishers, Boston. 50, 53, 63, 65, 66, 77, 78

Davari, S., Fazel Zarandi, M.H. & Hemmati, A. (2011). Maximal covering location problem (MCLP) with fuzzy travel times. *Expert Systems with Applications*, **38**, 14535–14541. 65

Department of Health (2010). National Department of Health Strategic Plan 2010/11 - 2012/13. 12, 13, 148, 154

Dimopoulou, M. & Giannikos, I. (2007). Advances in location analysis. *European Journal of Operational Research*, **179**, 923–926. 49

Ehrmeyer, S.S. & Laessig, R.H. (2007). Point-of-care testing, medical error, and patient safety: a 2007 assessment. *Clinical chemistry and laboratory medicine*, **45**, 766–73. 20

Elbireer, A.M., Opio, A.A., Brough, R.L., Jackson, J.B. & Manabe, Y.C. (2011). Strengthening public laboratory service in Sub-Saharan Africa: Uganda case study. *Laboratory Medicine*, **42**, 719–725. 18

Erdemir, E.T., Batta, R., Rogerson, P.A., Blatt, A. & Flanigan, M. (2010). Joint ground and air emergency medical services coverage models: A greedy heuristic solution approach. *European Journal of Operational Research*, **207**, 736–749. 57

Farahani, R.Z. & Masoud, H., eds. (2009). *Facility location: Concepts, models, algorithms and case studies*. Springer, Dordrecht. 73, 74

Farahani, R.Z., SteadieSeifi, M. & Asgari, N. (2010). Multiple criteria facility location problems: A survey. *Applied Mathematical Modelling*, **34**, 1689–1709. 81

Farahani, R.Z., Asgari, N., Heidari, N., Hosseininia, M. & Goh, M. (2012). Covering problems in facility location: A review. *Computers & Industrial Engineering*, **62**, 368–407. 51, 53, 64

Galvao, R.D., Gonzalo, L., Espejo, A. & Boffey, B. (2002). Discrete Optimization: A hierarchical model for the location of perinatal facilities in the municipality of Rio de Janeiro. *European Journal of Operational Research*, **138**, 495–517. 55

Gendreau, M., Laporte, G. & Semet, F. (2001). A dynamic model and parallel tabu search heuristic for real-time ambulance relocation. *Parallel Computing*, **27**, 1641–1653. 56

Geroliminis, N., Kepaptsoglou, K. & Karlaftis, M.G. (2011). A hybrid hypercube: Genetic algorithm approach for deploying many emergency response mobile units in an urban network. *European Journal of Operational Research*, **210**, 287–300. 57

Getahun, H., Harrington, M., O'Brien, R. & Nunn, P. (2007). Diagnosis of smear-negative pulmonary tuberculosis in people with HIV infection or AIDS in resource-constrained settings: Informing urgent policy changes. *Lancet*, **369**, 2042–9. 22

Ghaderi, A. & Jabalameli, M.S. (2013). Modeling the budget-constrained dynamic uncapacitated facility location-network design problem and solving it via two efficient heuristics: A case study of health care. *Mathematical and Computer Modelling*, **57**, 382–400. 56

Glencross, D., Coetzee, L. & Cassim, N. (2014). Integrated tiered CD4 service delivery in a national programme. *PloS ONE*, submitted. 102

Glencross, D.K., Janossy, G., Coetzee, L.M., Lawrie, D., Aggett, H.M., Scott, L.E., Sanne, I., McIntyre, J.a. & Stevens, W. (2008). Large-scale affordable PanLeucogated CD4+ testing with proactive internal and external quality assessment: In support of the South African national comprehensive care, treatment and management programme for HIV and AIDS. *Cytometry. Part B, Clinical Cytometry*, **74 Suppl 1**, S40–51. 32

Glencross, D.K., Coetzee, L.M., Faal, M., Masango, M., Stevens, W.S., Venter, W.F. & Osih, R. (2012). Performance evaluation of the Pima point-of-care CD4 analyser using capillary blood sampling in field tests in South Africa. *Journal of the International AIDS Society*, **15**, 3. 24, 32, 35, 36

Guo, M., Li, B., Zhang, Z., Wu, S. & Song, J. (2013). Efficiency evaluation for allocating community-based health services. *Computers & Industrial Engineering*, **65**, 395–401. 55

Hale, T.S. & Moberg, C.R. (2003). Location science research: A review. *Annals of Operations Research*, **123**, 21–35. 46, 47, 48, 49, 60, 62

Hauman, C. & Bekker, J. (2014). Application of the multi-objective cross-entropy method to the vehicle routing problem with soft time windows. *ORiON*, **30**, 19–40. 108

Health Systems Trust (2010). The 10 point plan. *Kwik Skwiz*, **2**, 1–2. 13

Health Systems Trust (2013). *South African Health Review 2012/13*. Durban. 23

Heller, T., Lessells, R.J., Wallrauch, C.G., Bärnighausen, T., Cooke, G.S., Mhlongo, L., Master, I. & Newell, M.L. (2010). Community-based treatment for multidrug-resistant tuberculosis in rural KwaZulu-Natal, South Africa. *The International Journal of Tuberculosis and Lung Disease*, **14**, 420–6. 23

Hillier, F.S. & Lieberman, G.J. (2010). *Introduction to Operations Research*. McGraw Hill, New York, ninth edn. 39, 41

Ho, W. (2008). Integrated analytic hierarchy process and its applications – A literature review. *European Journal of Operational Research*, **186**, 211–228. 51, 53

Human, A. (2010). A tale of two tiers: Inequality in South Africas health care system. *University of British Columbia Medical Journal*, **2**, 33. 9

International Conference of the African Society for Laboratory Medicine (2012). Ministerial call for action: Strengthening laboratory services in Africa. 20

Kim, D.G. & Kim, Y.D. (2010). A branch and bound algorithm for determining locations of long-term care facilities. *European Journal of Operational Research*, **206**, 168–177. 55

Knight, V., Harper, P. & Smith, L. (2012). Ambulance allocation for maximal survival with heterogeneous outcome measures. *Omega*, **40**, 918–926. 57

Landa-Torres, I., Manjarres, D., Salcedo-Sanz, S., Del Ser, J. & Gil-Lopez, S. (2013). A multi-objective grouping Harmony Search algorithm for the optimal distribution of 24-hour medical emergency units. *Expert Systems with Applications*, **40**, 2343–2349. 56

Larson, B., Schnippel, K., Ndibongo, B., Long, L., Fox, M.P. & Rosen, S. (2012). How to estimate the cost of point-of-care CD4 testing in program settings: an example using the Alere Pima Analyzer in South Africa. *PloS ONE*, **7**, e35444. 37

Larson, B.A., Brennan, A., McNamara, L., Long, L., Rosen, S., Sanne, I. & Fox, M.P. (2010). Lost opportunities to complete CD4+ lymphocyte testing among patients who tested positive for HIV in South Africa. *Bulletin of the World Health Organization*, **88**, 675–80. 29, 30

Lehe, J.D., Fuller, D., Schnippel, K., Glencross, D.K., Cassim, N. & Coetzee, L.M. (2014). Personal interview. 4 March, Johannesburg. 96

Lessells, R.J., Mutevedzi, P.C., Cooke, G.S. & Newell, M.L. (2011). Europe PMC funders group retention in HIV care for individuals not yet eligible for antiretroviral therapy: Rural KwaZulu-Natal, South Africa. *Journal of Acquired Immune Deficiency Syndrome*, **56**, 79–86. 26

Lessells, R.J., Cooke, G.S., McGrath, N., Nicol, M.P., Newell, M.L. & Godfrey-Faussett, P. (2013). Impact of a novel molecular TB diagnostic system in patients at high risk of TB mortality in rural South Africa (Uchwepheshe): Study protocol for a cluster randomised trial. *Trials*, **14**, 170. 24

Manabe, Y.C., Wang, Y., Elbireer, A., Auerbach, B. & Castelnuovo, B. (2012). Evaluation of portable point-of-care CD4 counter with high sensitivity for detecting patients eligible for antiretroviral therapy. *PloS ONE*, **7**, e34319. 35, 36

Maputo Conference (2008). Consultation on technical and operational recommendations for clinical laboratory testing harmonization and standardization. 18, 27

Mayosi, B.M., Lawn, J.E., van Niekerk, A., Bradshaw, D., Abdool Karim, S.S. & Coovadia, H.M. (2012). Health in South Africa: Changes and challenges since 2009. *Lancet*, **380**, 2029–43. 9, 10, 16, 23

Melo, M., Nickel, S. & Saldanha-da Gama, F. (2009). Facility location and supply chain management – A review. *European Journal of Operational Research*, **196**, 401–412. 51, 53

Meyer-Rath, G., Schnippel, K., Long, L., MacLeod, W., Sanne, I., Stevens, W., Pillay, S., Pillay, Y. & Rosen, S. (2012). The impact and cost of scaling up GeneXpert MTB/RIF in South Africa. *PloS ONE*, **7**, e36966. 22, 23

Mitropoulos, P., Mitropoulos, I. & Giannikos, I. (2013). Combining DEA with location analysis for the effective consolidation of services in the health sector. *Computers & Operations Research*, **40**, 2241–2250. 56

Mooney, G. & Gilson, L. (2009). The economic situation in South Africa and health inequities. *Lancet*, **374**, 858–9. 12

Motsoaledi, A. (2012). Progress and changes in the South African health sector. *Lancet*, **380**, 1969–70. 16

Mtapuri-Zinyowera, S., Chideme, M., Mangwanya, D., Mugurungi, O., Gudukeya, S., Hatzold, K., Mangwiro, A., Bhattacharya, G., Lehe, J. & Peter, T. (2010). Evaluation of the PIMA point-of-care CD4 analyzer in VCT clinics in Zimbabwe. *Journal of Acquired Immune Deficiency Syndrome*, **55**, 1–7. 35, 36

Mugglin, C., Estill, J., Wandeler, G., Bender, N., Egger, M., Gsponer, T. & Keiser, O. (2012). Loss to programme between HIV diagnosis and initiation of antiretroviral therapy in sub-Saharan Africa: Systematic review and meta-analysis. *Tropical Medicine & International Health*, **17**, 1509–20. 27, 28, 29, 30, 99

Murawski, L. & Church, R.L. (2009). Improving accessibility to rural health services: The maximal covering network improvement problem. *Socio-Economic Planning Sciences*, **43**, 102–110. 56

Mwau, M., Adungo, F., Kadima, S., Njagi, E., Kirwaye, C., Abubakr, N.S., Okubi, L.A., Waihenya, M., Lusike, J. & Hungu, J. (2013). Evaluation of PIMA point of care technology for CD4 T cell enumeration in Kenya. *PloS ONE*, **8**, e67612. 35, 36

Naoum-Sawaya, J. & Elhedhli, S. (2013). A stochastic optimization model for real-time ambulance redeployment. *Computers & Operations Research*, **40**, 1972–1978. 57

National Health Laboratory Service Website (2014). National Health Laboratory Service. Online, www.nhls.co.za, accessed on 17 January 2014. 18, 19

National Treasury (2013). Budget 2013: Estimates of national expenditure – Vote 16: Health. Tech. rep., National Treasury: Republic of South Africa, Pretoria. 24

Nkengasong, J.N. (2010). A shifting paradigm in strengthening laboratory health systems for global health: acting now, acting collectively, but acting differently. *American Journal of Clinical Pathology*, **134**, 359–60. 18

Nkengasong, J.N., Nsubuga, P., Nwanyanwu, O., Gershy-Damet, G.M., Roscigno, G., Bulterys, M., Schoub, B., DeCock, K.M. & Birx, D. (2010). Laboratory systems and services are critical in global health: Time to end the neglect? *American Journal of Clinical Pathology*, **134**, 368–73. 20

Olmsted, S.S., Moore, M., Meili, R.C., Duber, H.C., Wasserman, J., Sama, P., Mundell, B. & Hilborne, L.H. (2010). Strengthening laboratory systems in resource-limited settings. *American Journal of Clinical Pathology*, **134**, 374–80. 20

Owen, S.H. & Daskin, M.S. (1998). Strategic facility location: A review. *European Journal of Operational Research*, **111**, 423–447. 47, 48, 60, 61, 68, 69

Pai, N.P., Vadnais, C., Denkinger, C., Engel, N. & Pai, M. (2012). Point-of-care testing for infectious diseases: Diversity, complexity, and barriers in low- and middle-income countries. *PLoS Medicine*, **9**, e1001306. 22

Petti, C.a., Polage, C.R., Quinn, T.C., Ronald, A.R. & Sande, M.A. (2006). Laboratory medicine in Africa: A barrier to effective health care. *Clinical Infectious Diseases*, **42**, 377–82. 18, 19

Plate, D.K. (2007). Evaluation and implementation of rapid HIV tests: The experience in 11 African countries. *AIDS Research and Human Retroviruses*, **23**, 1491–8. 24

Reid, P.P., Compton, W.D., Grossman, J.H. & Fanjiang, G. (2005). *Building a better delivery system: A new engineering / health care partnership*. The National Academies Press, Washington. 43, 44

ReVelle, C. & Eiselt, H. (2005). Location analysis: A synthesis and survey. *European Journal of Operational Research*, **165**, 1–19. 46, 47, 48, 49, 50, 61, 62, 72

Rosen, S. & Fox, M.P. (2011). Retention in HIV care between testing and treatment in sub-Saharan Africa: A systematic review. *PLoS Medicine*, **8**, e1001056. 99, 100

Rubinstein, R.Y. (1997). Optimization of computer simulation models with rare events. *European Journal of Operational Research*, **99**, 89–112. 108

Rubinstein, R.Y. & Kroese, D.P. (2004). *The cross-entropy method: A unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning*. Springer. 108

Ruff, B., Mzimba, M., Hendrie, S. & Broomberg, J. (2011). Reflections on health-care reforms in South Africa. *Journal of Public Health Policy*, **32**, S184–S192. 12

Schmid, V. (2012). Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. *European Journal of Operational Research*, **219**, 611–621. 57

Scholtz, E. (2014). *A comparative study on the value of accounting for possible relationships between decision variables when solving multi-objective problems*. Master's thesis, Stellenbosch University. 108

Sha, Y. & Huang, J. (2012). The multi-period location-allocation problem of engineering emergency blood supply systems. *Systems Engineering Procedia*, **5**, 21–28. 56

Shariat-Mohaymany, A., Babaei, M., Moadi, S. & Amiripour, S.M. (2012). Linear upper-bound unavailability set covering models for locating ambulances: Application to Tehran rural roads. *European Journal of Operational Research*, **221**, 263–272. 57

Shariff, S.R., Moin, N.H. & Omar, M. (2012). Location allocation modeling for healthcare facility planning in Malaysia. *Computers & Industrial Engineering*, **62**, 1000–1010. 55

Smith, H.K., Harper, P.R., Potts, C.N. & Thyle, A. (2009). Planning sustainable community health schemes in rural areas of developing countries. *European Journal of Operational Research*, **193**, 768–777. 55

South African National AIDS Council (2012). National Strategic Plan For HIV, STIs and TB 2012 - 2016. Tech. rep., The South African National AIDS Council. 14, 15

South African Press Agency (2013). 84% of South Africans get 2nd rate healthcare – Motsoaledi. Online, www.news24.com, published on 12 September 2013. 9

Stadler, J.G. (2012). *Multi-objective optimisation using the cross- entropy method in CO gas management at a South African ilmenite smelter*. Master's thesis, Stellenbosch University. 108

Stevens, W.S., Glencross, D.K., Cassim, N. & Coetzee, L.M. (2014). Personal interview. 12 June, Johannesburg. 95, 96, 107

Sukapirom, K., Onlamoon, N., Thepthai, C., Polsrila, K., Tassaneetrithep, B. & Pattanapanyasat, K. (2011). Performance evaluation of the Alere PIMA CD4 test for monitoring HIV-infected individuals in resource-constrained settings. *Journal of Acquired Immune Deficiency Syndromes*, **58**, 141–7. 35, 36

The Presidency: Republic of South Africa (2009). Medium term strategic framework. 12

UNAIDS (2011). *World AIDS Day Report 2011*. Joint United Nations Programme on HIV/AIDS (UNAIDS). 15, 16

UNAIDS (2013). *Global report: UNAIDS report on the global AIDS epidemic 2013*. Joint United Nations Programme on HIV/AIDS (UNAIDS). 15, 16, 23

UNAIDS (2014). AIDSinfo Online Database. Online, www.aidsinfoonline.org, accessed on 13 March 2014. 17

Van Rie, A., Page-Shipp, L., Scott, L., Sanne, I. & Stevens, W. (2010). Xpert MTB/RIF for point-of-care diagnosis of TB in high-HIV burden, resource-limited countries: Hype or hope? *Expert Review of Molecular Diagnostics*, **10**, 937–46. 22, 23

Winston, W.L. (2004). *Operations Research: Applications and algorithms*. Brooks/-Cole, Belmont, 4th edn. 39, 40

World Bank (2013a). World Developement Indicators: Healthcare expenditure as a percentage of GDP. 11

World Bank (2013b). World Development Indicators: Health systems. Tech. rep., The World Bank, Washington DC. 10

World Bank (2013c). World Development Indicators: Public health expenditure as a percentage of total health expenditure. 11

World Health Organisation (2013a). *Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection*. June, WHO Press, Geneva. 26

World Health Organisation (2013b). *HIV assays: Operational characteristics*. WHO Press, Geneva. 24

World Health Organisation (2014). World Health Organisation Diagnostic and Laboratory Technology. Online, www.who.int/diagnostics, accessed on 10 March 2014. 24

Wu, G. & Zaman, M.H. (2012). Low-cost tools for diagnosing and monitoring HIV infection in low-resource settings. *Bulletin of the World Health Organization*, **90**, 914–20. 24, 25, 32, 34, 35

Wynberg, E., Cooke, G., Shroufi, A., Reid, S.D. & Ford, N. (2014). Impact of point-of-care CD4 testing on linkage to HIV care: A systematic review. *Journal of International AIDS Society*, **17**, 1–6. 3, 25, 31, 93, 99, 100

Zachariah, R., Reid, S.D., Chaillet, P., Massaquoi, M., Schouten, E.J. & Harries, A.D. (2011). Viewpoint: Why do we need a point-of-care CD4 test for low-income countries? *Tropical Medicine & International Health*, **16**, 37–41. 30, 32, 34, 35

Appendix A

South African healthcare sector strategic goals and objectives

This appendix contains more detailed information on the South African public healthcare sector's strategic goals and objectives. The South African government's Medium Term Strategic Framework (MTSF) as well as the National Department of Health's set of priorities (the "10 Point Plan"), drawn up in response to the MTSF, were introduced in section 2.1.4. Here the details of how the 10 Point Plan relates to the 20 healthcare outcomes defined in the MTSF are given. The 2014 target for each of the MTSF healthcare outcomes is also defined.

A.1 The 10 Point Plan and the Medium Term Strategic Framework

In its strategic plan for the period 2010/11 to 2012/13, the National Department of Health gives a table that demonstrates how the 10 Point Plan is aligned to the 20 healthcare outcomes listed in the MTSF. This table also outlines the key activities to be executed for each of the priorities in the 10 Point Plan. This table has been reproduced here.

Table A.1: The 10 Point Plan and the Medium Term Strategic Framework. (Reproduced from: Department of Health (2010).)

Department of Health 10 Point Plan 2009-2014		**Deliverables from the outcome-based MTSF 2009 -2014**
Priorities	**Key activities**	
1. Provision of Strategic leadership and creation of Social compact for better health outcomes	Ensure unified action across the health sector in pursuit of common goals. Mobilize leadership structures of society and communities. Communicate to promote policy and buy in to support government programs. Review of policies to achieve goals. Impact assessment and program evaluation. Development of a social compact. Grassroots mobilization campaign.	Revitalisation of the Primary Health Care approach. Enhanced Operational Management of Health Facilities.
2. Implementation of National Health Insurance (NHI)	Finalisation of NHI policies and implementation plan. Immediate implementation of steps to prepare for the introduction of the NHI, e.g. Budgeting, Initiation of the drafting of legislation. Finalise and implement an Information and Communication Technology (ICT) Strategy.	Improved Health Care Financing. Implementation of NHI. Strengthened Health Information Systems (HIS).

Continued on next page

Continued from previous page

Department of Health 10 Point Plan 2009-2014		**Deliverables from the outcome-based MTSF 2009 -2014**
Priorities	**Key activities**	
3. Improving the Quality of Health Services	Improve service delivery in all 52 districts, with a special emphasis on 18 priority districts. Refine and scale up the detailed plan on the improvement of Quality of services and directing its immediate implementation. Consolidate and expand the implementation of the Health Facilities Improvement Plans. Establish a National Quality Management and Accreditation Body.	Improved Patient Care and Satisfaction. Accreditation of health facilities for quality.
4. Overhauling the health care system and improving its management		
4.1 Refocus the Health System on Primary Health Care (PHC)	Develop and implement a national model for the delivery of health services based on the PHC approach. Scale up community-based promotive and preventive health service, and massively expand immunisation programmes: antenatal care post.	Revitalisation of the Primary Health Care approach. Enhanced Operational Management of Health Facilities.
4.2 Improve the functionality and management of the Health System	Assess the qualification, skills and competencies of Hospital CEOs; Hospital Senior Managers and District Managers. Training managers in leadership, management and governance. Decentralization of management. Development and implementation of an accountability framework for the public and private sectors. Establish a management and leadership	Revitalisation of the Primary Health Care approach. Enhanced Operational Management of Health Facilities.

Continued on next page

Continued from previous page

Department of Health 10 Point Plan 2009-2014		**Deliverables from the outcome-based MTSF 2009 -2014**
Priorities	**Key activities**	
	academy for health managers.	
5. Improved Human Resources Planning, Development and Management	Refinement of the HR plan for health. Re-opening of nursing schools and colleges. Recruitment and retention of professionals, including urgent collaboration with countries that have excess of these professionals. Focus on training of PHC personnel and midlevel health workers. Make an assessment of and also review the role of the Health Professional Training and Development Grant (HPTDG) and the National Tertiary Services Grant (NTSG). Manage the coherent integration and standardisation of all categories of Community Health Workers.	Improved access to Human Resources for Health.

Continued on next page

Continued from previous page

Department of Health 10 Point Plan 2009-2014		**Deliverables from the outcome-based MTSF 2009 -2014**
Priorities	**Key activities**	
6. Revitalization of physical infrastructure 6.1 Accelerate the delivery of health infrastructure through Public Private Partnerships (PPPs) 6.2 Revitalise Primary level facilities 6.3 Accelerate the delivery of Health Technology and Information Communication Technology (ICT) Infrastructure	 Establish Public Private Partnerships, particularly for the construction and refurbishment of Tertiary Hospitals. Accept 13 new projects annually for delivery through the revised Hospital Revitalisation Project. Implement refurbishment and preventative maintenance of all hospitals. Complete the Audit of PHC infrastructure and services. Accelerate the delivery of infrastructure for primary level facilities. Implement refurbishment and preventative maintenance of all hospitals. Finalise and implement the Health Technology Strategy. Finalise and implement the ICT Strategy for the Health Sector.	Improved Physical Infrastructure for Healthcare Delivery.
7. Accelerated implementation of the HIV and AIDS strategic plan and the increased focus on TB and other communicable diseases	Implement new HIV and AIDS policies and strategies announced on World AIDS Day, 01 December 2009. Urgently strengthen programs against TB, MDR-TB and XDR-TB. Implement new PMTCT Guidelines.	Managing HIV Prevalence. Reduced HIV Incidence. Expanded PMTCT Programme. Improved TB Case Finding. Improved TB outcomes. Improved access to Antiretroviral

Continued on next page

Continued from previous page

Department of Health 10 Point Plan 2009-2014		**Deliverables from the outcome-based MTSF 2009 -2014**
Priorities	**Key activities**	
		Treatment for HIV-TB co-infected patients. Decreased prevalence of MDR-TB. Expanded access to Home Based Care and Community Health Workers.
8. Mass mobilisation for the better health for the population	Place more focus on the programs to attain the Millennium Development Goals (MDGs). Intensify health promotion programs. Place more focus on Maternal, Child and Womens Health. Place more focus on non-communicable diseases and patients rights, quality and provide accountability.	Increased Life Expectancy at Birth. Reduced Child Mortality. Decreased Maternal Mortality Ratio. Improved health services for the Youth. Expanded access to Home Based Care and Community Health Workers.
9. Review of drug policy:	Complete and submit proposals and a strategy, with the involvement of various stakeholders. Draft plans for the establishment of a Stateowned drug manufacturing entity.	Improved Patient Care and Satisfaction. Accreditation of health facilities for quality. Enhanced Operational Management of Health Facilities.
10. Strengthen Research and Development	Commission research to accurately quantify Infant mortality. Commission research into the impact of social determinants of health and nutrition. Support research studies to promote indigenous knowledge systems and the use of appropriate traditional medicines.	Enhanced Operational Management of Health Facilities. Strengthened Health Information Systems (HIS).

A.2 Quantified 2009 performance and 2014 target for the MTSF healthcare outcomes

In its strategic plan for the period 2010/11 to 2012/13, the National Department of Health gives a table that:

1. Quantifies the country's performance in terms of each of the 20 healthcare outcomes listed in the MTSF; and

2. More clearly defines the target for each of these 20 healthcare outcomes.

This table is reproduced here.

Table A.2: 2009 performance and 2014 target for the MTSF healthcare outcomes. (Reproduced from: Department of Health (2010).)

Indicator	**Baseline 2009**	**Target 2014/2015**
Life Expectancy at Birth	53.9 years for males. 57,2 years for females.	58-60 years.
Child Mortality	69 per 1,000 live births.	30 - 45 per 1,000 live births.
Maternal Mortality Ratio	400 - 625 per 100,000 live births.	100 per 100,00 live births.
HIV Prevalence (amongst 15-24 year old pregnant women)	21,7%	Not Applicable.
HIV Incidence	1,3%	0,6%
Mother to child transmission rate of HIV	10%	0% - ¡ 5%
Percentage of eligible HIV positive women initiated on ART	37%	All eligible pregnant women to be initiated on ART at a CD4 count of ¡350 or WHO stage III or IV.
TB cases notified	341, 165	175,000
TB Cure Rate	64%	85%
Percentage of HIV-TB co-infected patients who are on ART	30%	100%
PHC service delivery model completed	Strategy for accelerating progress towards health related MDGs through strengthening PHC developed.	Health service delivery model based on the PHC approach developed.
Percentage of health facilities accredited for quality	None.	25% of health facilities accredited annually.

Continued on next page

Continued from previous page

Indicator	**Baseline 2009**	**Target 2014/2015**
Improved access to Human Resources for Health	Human Resources for Health (HRH) Plan produced.	Revised HRH Plan produced, which reflects an appropriate balance between health professionals and administrative personnel; reintroduces key PHC workers such as Infection Control Officers; Environmental Health Practitioners. Monitor vacancy rates in the public sector on a quarterly basis.
Improved Health Care Financing	Creation of national Health Insurance comenced.	NHI policy finalised and implemented.
Strengthened Health information systems (HIS)	Draft e-Health Strategy produced. National Indicator Dataset (NIDS) revised.	Finalise e-Health Strategy finalised and implemented. Finalise new NIDS.
Improved health services for the Youth	Strategy to improve health levels of the youth population segments developed.	Strategy finalised. 70% of PHC facilities implementing Youth Friendly Services by 2014/15.
Expanded access to Home Based Care and Community Health Workers	Draft policy on Community Health Workers produced.	Strategy for Home and Community-based Care (HCBC) developed. Policy on Community Health Workers finalised.