University of Cape Town

Resource constraints in an epidemic: A goal programming and mathematical modelling framework for optimal resource shifting in South Africa

Student:
Saadiyah Mayet
MYTSAA002

Supervisor:
A/Prof Sheetal Silal
Co-supervisor:
A/Prof Ian Durbach

Dissertation presented in partial fulfillment of the degree

M.Sc. Advanced Analytics

Department of Statistical Sciences

October 13, 2021

Declaration of Authorship

Signed: Signed by candidate

Date: 21 October 2021

Abstract

The COVID-19 pandemic has had devastating consequences across the globe, and has led many governments into completely new decision making territory. Developing models which are capable of producing realistic projections of disease spread under extreme uncertainty has been paramount for supporting decision making by many levels of government. In South Africa, this role has been fulfilled by the South African COVID-19 Modelling Consortium's generalised Susceptible-Exposed-Infectious-Removed compartmental model, known as the National COVID-19 Epi Model. This thesis adapted and contributed to the Model; its primary contribution has been to incorporate the feature that resources available to the health system are limited. Building capacity constraints into the Model allowed it to be used in the resource-scarce context of a pandemic. This thesis further designed and implemented a goal programming framework to shift ICU beds between districts intra-provincially in a way that aimed to minimise deaths caused by the non-availability of ICU beds. The results showed a 15% to 99% decrease in lives lost when ICU beds were shifted, depending on the scenario considered. Although there are limitations to the scope and assumptions of this thesis, it demonstrates that it is possible to combine mathematical modelling with optimisation in a way that may save lives through optimal resource allocation.

Acknowledgements

I would first and foremost like to thank Associate Professor Sheetal Silal for being an exemplary supervisor and mentor. Not only has she guided me through this thesis, but she inspires me everyday with her dedication to public health, and has taught me the importance of compassion, co-operation, and a strong conscience in research. I would similarly like to thank Associate Professor Ian Durbach, for complementing Sheetal's supervision in the best manner possible, and for always pushing me to think in smarter ways.

I would like to acknowledge the support of several people outside of my supervision team: Dr Gesine Meyer-Rath and Lise Jamieson from the Health Economics and Epidemiology Research Office, who provided the district-wise hospital bed allocation dataset that was central to this thesis; Nikhil Khanna and Tucker Bbosa from the Clinton Health Access Initiative, who guided my thinking and assisted in providing provincial ambulance data; and Thomas Crompton from the Right to Care organisation, who provided invaluable insights on the COVID-19 context in South Africa, particularly with regards to hospital planning.

I would like to thank my family: my parents, Ebrahim and Shenaaz, whose unwavering support and pride kept me going through the most difficult parts of this thesis; and my sisters, Adhila and Naseeha, for being my biggest fans and my first role models.

I would further like to acknowledge the role played by all of my friends throughout my Master's degree. They have made an otherwise daunting task one which has been filled with laughter and unexpected learning.

Finally, I would like to acknowledge my original sounding board, Sassy the cat, whose presence has been sorely missed.

Contents

List of Figures

List of Tables

List of Abbreviations

ABM	**A**gent **B**ased **M**odel
CHAI	**C**linton **H**ealth **A**ccess **I**nitiative
COVID-19	**Co**rona**v**irus **D**isease 2019
DES	**D**iscrete **E**vent **S**imulation
GHS	**G**eneral **H**ousehold **S**urvey
HE2RO	**H**ealth **E**conomics (and) **E**pidemiology **R**esearch **O**ffice
ICU	**I**ntensive **C**are **U**nit
LMIC	**L**ow (and) **M**iddle **I**ncome **C**ountry
LOS	**L**ength **O**f **S**tay
NCEM	**N**ational **C**OVID-19 **E**pi **M**odel
NDOH	**N**ational **D**epartment **O**f **H**ealth
NPI	**N**on-**P**harmaceutical **I**ntervention
OR	**O**perations **R**esearch
PPE	**P**ersonal **P**rotective **E**quipment
RSA	**R**epublic (of) **S**outh **A**frica
SA	**S**outh **A**frica
SACMC	**S**outh **A**frican **C**OVID-19 **M**odelling **C**onsortium
SEIR	**S**usceptible-**E**xposed-**I**nfectious-**R**ecovered
UK	**U**nited **K**ingdom
USA	**U**nited **S**tates (of) **A**merica

DISTRICTS

Eastern Cape:

DC10	Sarah Baartman District Municipality
DC12	Amathole District Municipality
DC13	Chris Hani District Municipality
DC14	Joe Gqabi District Municipality
DC15	OR Tambo District Municipality
DC44	Alfred Nzo District Municipality
BUF	Buffalo City Metropolitan Municipality
NMA	Nelson Mandela Bay Metropolitan Municipality

Free State:

DC16	Xhariep District Municipality
DC18	Lejweleputswa District Municipality
DC19	Thabo Mofutsanyana District Municipality
DC20	Fezile Dabi District Municipality
MAN	Mangaung Metropolitan Municipality

Gauteng:

DC42	Sedibeng District Municipality
DC48	West Rand District Municipality

EKU	Ekurhuleni Metropolitan Municipality
JHB	City of Johannesburg Metropolitan Municipality
TSH	City of Tshwane Metropolitan Municipality

KwaZulu-Natal:

DC21	Ugu District Municipality
DC22	uMgungundlovu District Municipality
DC23	uThukela District Municipality
DC24	uMzinyathi District Municipality
DC25	Amajuba District Municipality
DC26	Zululand District Municipality
DC27	uMkhanyakude District Municipality
DC28	King Cetshwayo District Municipality
DC29	iLembe District Municipality
DC43	Harry Gwala District Municipality
ETH	eThekwini Metropolitan Municipality

Limpopo:

DC33	Mopani District Municipality
DC34	Vhembe District Municipality
DC35	Capricorn District Municipality
DC36	Waterberg District Municipality
DC47	Sekhukhune District Municipality

Mpumalanga:

DC30	Gert Sibande District Municipality
DC31	Nkangala District Municipality
DC32	Ehlanzeni District Municipality

North West:

DC37	Bojanala Platinum District Municipality
DC38	Ngaka Modiri Molema District Municipality
DC39	Dr Ruth Segomotsi Mompati District Municipality
DC40	Dr Kenneth Kaunda District Municipality

Northern Cape:

DC6	Namakwa District Municipality
DC7	Pixley ka Seme District Municipality
DC8	Mgcawu District Municipality
DC9	Frances Baard District Municipality
DC45	John Taolo Gaetsewe District Municipality

Western Cape:

DC1	West Coast District Municipality
DC2	Cape Winelands District Municipality
DC3	Overberg District Municipality
DC4	Garden Route District Municipality
DC5	Central Karoo District Municipality
CPT	City of Cape Town Metropolitan Municipality

For my uncle, Yusuf Kaloo, who would have been so proud.

1 Introduction

This chapter will provide brief context on the effects of COVID-19 and the role of infectious disease modelling in guiding governmental decision making. It will further outline the research objectives and scope of this thesis.

1.1 Background

The COVID-19 pandemic has burdened health systems to the point of collapse globally, and health resource scarcity has undoubtedly resulted in higher mortality than it would have if nations were operating under unlimited health system capacity. This holds especially true in regions like Africa which face, for example, dire Intensive Care Unit (ICU) bed shortages under normal circumstances (Dzinamarira, Dzobo, and Chitungo, 2020). At the time of writing, there have been over 46 million reported cases globally, and 725 452 in South Africa (SA) in particular (Mkhize, 2020; Johns Hopkins University and Medicine, 2021).

Both the developed and developing worlds have struggled under the strain of COVID-19. Some examples include when Italy's world-class healthcare system buckled under the stress of the pandemic, with hospitals running out of beds and staff (Chow and Saliba, 2020), while in Houston, Texas, ICU beds ran out to the point that COVID-19 patients had to be treated in sub-optimal emergency rooms. The pandemic also saw London's Northwick Park hospital running short of ICU beds, resulting in treatment of COVID-19 patients in operating theatres (Campbell, Marsh, and Bannock, 2020). In South Africa, many health facilities have suffered shortages brought on or exacerbated by the pandemic. Not only have hospitals faced staff shortages due to coronavirus infections, but the staff of many facilities have gone on strike due to understaffing and a lack of personal protective equipment (PPE), especially in the Eastern Cape (Lepule, 2020; SA Provincial Health, 2020; Harding, 2020). Major Western Cape hospital Tygerberg ran out of ICU beds, while Gauteng faced oxygen shortages (Haffajee, 2020; Petersen, 2020). These widespread health system failings indicate a need for better planning, and for better sharing of resources between resource-rich and resource-scarce regions.

In the face of such scarcity, governments around the world have found themselves in largely uncharted decision making territory, and as a result, the importance of reliable decision support under uncertainty has never been more apparent. Mathematical

modelling of the spread of COVID-19 has both led public thinking around intervention measures, and come under intense scrutiny, the most obvious example being Imperial College London's simulation models from March of 2020 (Adam, 2020). The results of these compartmental models were shared both with the United Kingdom (UK) government, and the White House in the United States of America (USA), and are said to have influenced action in both countries (Adam, 2020).

In South Africa, a multidisciplinary group known as the South African COVID-19 Modelling Consortium (SACMC) was formed to help guide decision makers on various levels of government throughout the course of the COVID-19 pandemic. The group, of which the candidate is a member, developed and implemented a population-level compartmental model known as the National COVID-19 Epi Model (NCEM) to project disease spread on a district level in South Africa. However, the resource-scarce context of a pandemic specifically calls for models which incorporate shortages and can thereby examine their effects on mortality. Moreover, there is a need for mitigation of these effects through intelligent allocation of resources, which can be achieved through an optimisation framework.

1.2 Research Objectives

In the development of the NCEM at the start of the epidemic in South Africa, this research contributed the inclusion of general and ICU bed capacity constraints (COVID-19 Modelling Consortium, 2020). Furthermore, this thesis built the epidemiological model into a wider optimisation framework, which uses a goal programming approach to shift resources amongst districts in a way that is optimal for reducing shortages, and in turn minimising loss of life. The research objectives can be summarised as follows:

1. Provide a descriptive analysis of the data for this thesis.
2. Adapt the SACMC's model to be able to measure excess mortality due to the non-availability of ICU and general hospital beds in each district.
   (a) Implement and examine the results of these capacity constraints.
   (b) Examine excess mortality in each of the districts.
   (c) Explore the relationship between the availability of resources and excess mortality.
3. Develop a goal programming framework which aims to minimise excess mortality by optimally shifting resources amongst districts over the course of a year.
   (a) Design a number of scenarios under which to implement the goal programming and simulation framework.
   (b) Explore under which circumstances (if any) the reduction in excess mortality is greatest.

1.3 Scope

Global scientific knowledge of COVID-19 is growing on a daily basis and, with the advent of second waves, the global context of the pandemic continues to develop

rapidly. The research for this thesis began in March 2020, and the literature upon which the analysis is based was published primarily by the first half of 2020. Where possible, as new information became available, it was included in this thesis; however, it is likely that some of the literature is outdated by the time this thesis is submitted for examination. While the candidate's research with the SACMC continues at a fast pace, the scientific knowledge of COVID-19 included in this thesis has been limited to what was published by October 2020. Similarly, the research for this thesis was in its final stages when the lineage B.1.351 emerged and brought about a second wave of COVID-19 in South Africa. Hence, this thesis focused on the first wave.

2

Literature Review

This chapter will lay down the context of COVID-19 and the South African health system. It will then explore the literature on the modelling of COVID-19 primarily in Africa and South Africa, with specific attention given to previous works which address the idea of limited capacity. It will further investigate the application of goal programming in healthcare resource allocation, and finally the combination of simulation and goal programming in healthcare, in order to paint a holistic picture of the literary foundation on which this thesis is built.

2.1 The COVID-19 context

Understanding how to tackle the issue of limited and spatially-variable capacity to fight COVID-19 requires first establishing an understanding of the disease itself, and the way in which it has affected, and been handled by, South Africa. COVID-19 is a respiratory disease caused by the coronavirus SARS-CoV-2, which originated in the Hubei province of China in December 2019 (The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, 2020). Respiratory droplets are the main mode of transmission of the disease, primarily via person-to-person contact, but secondarily through the contamination of surfaces (Wiersinga et al., 2020). One of the greatest challenges in containing the disease lies in the fact that presymptomatic carriers are infectious; in fact, it is estimated that around half of all transmission occurs in this way (Wiersinga et al., 2020). This is no surprise; as shown by He et al. (2020), infectiousness is highest in individuals before or at the time of symptom onset.

As demonstrated in a study in China, COVID-19 causes only mild symptoms in approximately 80% of cases (The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, 2020). Children tend to present with especially mild symptoms that are concentrated in the upper respiratory tract (Wiersinga et al., 2020). However, elderly people and individuals with co-morbidities are at higher risk of severe or critical infection (Sanyaolu et al., 2020). The disease has consequently resulted in more than a million fatalities worldwide at the time of writing (Johns Hopkins University and Medicine, 2021).

While as of November 2020 there is no cure for COVID-19, various pharmaceutical and non-pharmaceutical interventions have been successfully used to combat the disease. Chu et al. (2020) showed that maintaining a distance of at least one metre between

people and wearing face masks both significantly reduce the chances of transmission. Additionally, hand washing, contact tracing and limiting the size of gatherings all help to contain the spread of the disease (Wiersinga et al., 2020). There are two drugs which have largely helped curb the damage caused by COVID-19. The antiviral drug Remdesivir helps speed up the time to recovery (Beigel et al., 2020), while the anti-inflammatory Dexamethasone has been shown to reduce mortality amongst patients on oxygen or mechanical ventilation (Horby et al., 2020). Most hospitalised patients require oxygenation; when conventional oxygenation fails, high-flow nasal canula oxygen is administered (Wiersinga et al., 2020; Alhazzani et al., 2020). While mechanical ventilation can be necessary, unwarranted intubation can cause further complications (Wiersinga et al., 2020).

There are as yet many unknowns, not least of which is the strength and duration of immunity to the disease after infection. In October 2020, Ripperger et al. (2020) found in a study that antibodies were detectable for five to seven months after infection. However, whether this guarantees immunity to reinfection is unknown. Moreover, there has been evidence of reinfection around the world, albeit on a small scale (Tillett et al., 2020; Iwasaki, 2020). There has been no evidence of reinfection in South Africa (Ismail, 2020). At the time of writing, there are 36 vaccines in Phase 1 trials, 16 in Phase 2, 11 in Phase 3, and none approved for use (Kommenda and Hulley-Jones, 2020).

2.2 The South African context

Most of South Africa's over 59 million-strong population is served by the country's public health system (Statistics South Africa, 2020). According to the 2018 General Household Survey (GHS), 71.5% of households cited public health institutions, such as hospitals and clinics, as their first port of call when seeking healthcare (Statistics South Africa, 2018). Only 27.1% of households reported seeking care via private avenues, be it doctors' surgeries or private hospitals (Statistics South Africa, 2018). Health insurance or medical aid schemes exist purely for those who can afford them, and as of 2018, only 16.4% of the population were estimated to be protected by such schemes (Statistics South Africa, 2018). This means that the COVID-19 pandemic promised to place a large amount of pressure on an already strained public health system.

The provincial governments run primary, secondary, and tertiary public health institutions, while public policy formation falls in the hands of the National Ministry of Health (Mahlathi and Dlamini, 2015). As a result, the quality of and resources available to public healthcare tends to vary by province; for example, accessible healthcare is better in the Western Cape and Gauteng than it is in the Eastern Cape and Limpopo (Delobelle, 2013). This highlights the regional variation in ability to tackle a pandemic, which in turn supports the need for a spatial system of resource sharing. The choice of districts for the level on which resources should be shifted is further substantiated by the fact that most individuals access healthcare through the District Health System (Mahlathi and Dlamini, 2015).

South Africa's response to COVID-19 has been relatively stringent. Social distancing, school closure, and limitations to gatherings were enforced from the 18^{th} of March 2020 (RSA Government, 2020a), only six days after the first confirmed case of local transmission (Carlitz and Makhura, 2021). As of the 27^{th} of March, South Africa was subject to a "hard lockdown", in which people were required to remain at home unless essential workers, or unless travelling to and from supermarkets and pharmacies (RSA Government, 2020b). The lockdown was subsequently extended to the end of April (RSA Government, 2020c), after which the country was moved to Alert Level 4 (RSA Government, 2020e). The five Alert Levels were announced to be a reflection of both the level of disease spread, and health system readiness, with Level 5 representing "hard lockdown" and Level 1 indicating almost completely normal functioning of the country. South Africa moved to Alert Levels 3, 2, and 1 on the 1^{st} of June, 18^{th} of August, and the 21^{st} of September 2020 respectively (RSA Government, 2020f; RSA Government, 2020g; RSA Government, 2020h).

2.3 SEIR modelling of COVID-19

With context established around both the disease and the country at hand, the next step is to explore the historical use of Susceptible-Exposed-Infectious-Removed (SEIR) models in modelling COVID-19. There is extensive evidence of this in the literature (Kucharski et al., 2020; Davies et al., 2020; Pei and Shaman, 2020; Leung et al., 2020; Peng et al., 2020; Brand et al., 2020). This review will focus primarily on such modelling as applied to the African and South African contexts.

The spatial granularity of the generalised SEIR models applied to Africa tends to be very coarse. Davies et al. (2020) modelled Niger, Nigeria, and Mauritius on a national level, and acknowledged that models should separate rural and urban areas in order to account for slower and faster peaking respectively. Diop et al. (2020) similarly modelled Ghana, Kenya, and Senegal on a national level, and also mentioned a non-ideal aggregation of rural and urban areas. All South African examples found modelled the country on a national level (Mukandavire et al., 2020; Zhao et al., 2020; Mushayabasa, Ngarakana-Gwasira, and Mushanyu, 2020; Nyabadza et al., 2020). Brand et al. (2020) modelled Kenya in twenty distinct regions, determined by population density, and then extrapolated results to the level of Kenya's 47 counties. The literature clearly indicates a need for a finer spatial granularity in the modelling of COVID-19 in Africa and South Africa.

Incorporating the effects of population density on the spread of COVID-19 is an intuitive aspect of COVID-19 modelling that is scarcely implemented in the literature. While Pei and Shaman (2020) scaled the force of infection - or rate at which susceptible individuals become exposed to COVID-19 - by population density, few other papers reviewed did more than comment on density as an aspect that should ideally be included. Brand et al. (2020) modelled on the level of patches divided according to density. In a study investigating the effects of social distancing on the spread of COVID-19, Nyabadza et al. (2020) commented on the potential for population density to affect ability to social distance, but did not account for it. The literature indicates both a precedent and a need for incorporating the effects of population density.

The purpose of modelling the spread of COVID-19 in Africa seems to exclude the issue of hospital resource demand and supply. Most studies are geared towards predicting general disease burden (Brand et al., 2020; Mukandavire et al., 2020) or assessing the impact of non-pharmaceutical interventions (NPIs) (Zhao et al., 2020; Taboe et al., 2020; Mushayabasa, Ngarakana-Gwasira, and Mushanyu, 2020; Nyabadza et al., 2020). Davies et al. (2020) modelled with the aim of quantifying the effect of NPIs, but specifically measured bed demand under each scenario given. Diop et al. (2020), meanwhile, refrained from estimating mortality due to an acknowledgement that capacity data is crucial for this, but was unavailable.

The only study reviewed which directly attempted to explore the event of capacity's running out in an African context was that of Barasa, Ouma, and Okiro (2020), which used the modelling estimates from another separate study (Brand et al., 2020). Barasa, Ouma, and Okiro (2020) estimated surge capacity and the point at which each of Kenya's counties would run out of hospital and ICU beds, but this estimation sequentially followed the modelling step rather than being integrated into it. Their approach also omitted the dynamics of hospital resources, such as the availability of trained healthcare workers (Barasa, Ouma, and Okiro, 2020). This demonstrates that the modelling literature scarcely examines hospital demand and supply in the African context.

Outside of Africa, the use of modelling to address the issue of limited hospital capacity also seems to be rare. Branas et al. (2020) directly sought to estimate excess mortality due to the non-availability of critical care beds. They based their analysis on the results of a generalised SEIR model created by Pei and Shaman (2020), which modelled disease spread in the USA on the level of the 3108 counties. They used a standard four compartment model. Branas et al. (2020) used the typical length of stay (LOS) of COVID-19 patients to calculate discharge rates, and retrospectively to calculate the number of patients who would have been turned away each day. These patients were assumed to have a mortality rate of 95%. They found that the highest such mortality was in urban counties. However, no account was given to the availability of healthcare workers and other hospital resources, such as ventilators. There is thus universal room for improvement in the analysis of excess mortality.

In South Africa, the multidisciplinary SACMC, of which the candidate is a member, was formed in late March 2020. The group developed the NCEM, a generalised SEIR stochastic transmission model, which was created and continually updated with the purpose of supporting decision making. Model results were presented to various stakeholders on an ongoing basis throughout the first wave of the COVID-19 epidemic in SA, such as the National Department of Health (NDOH), National Treasury, and provincial departments of health. The district-level model includes compartments for levels of disease severity - asymptomatic, pre-symptomatic, mild, severe, and critical - as well as for treatment pathways including general hospital admission, ICU care, and ventilation. This thesis is thus part of a wider project, and contributed the notion of limited general and ICU bed capacity to the NCEM (SA COVID-19 Modelling Consortium, 2020b). Note that the NCEM was not the only model reported in the media during the early stages of the COVID-19 pandemic in South Africa. Several others, including those by the Actuarial Society of South Africa, the firm Deloitte, and the Pandemic Data Analysis group (Low and Geffen, 2020) were producing results. The

candidate elected to work with the NCEM as its results were updated throughout the first wave of COVID-19.

2.4 Goal programming applied to healthcare

The most widespread application of goal programming in healthcare appears to be nurse scheduling (Azaiez and Sharif, 2005; Jenal et al., 2011; Agyei, Obeng-Denteh, and Andaam, 2015; Ferland, Ahiod, and Michelon, 2001; Gür and Eren, 2018; Huarng, 1999). Consistent with the findings of Moon et al. (1989), health resource allocation goal programming problems in the literature are more often applied within a particular institution than they are on a large scale (Arenas et al., 2002; Ataollahi et al., 2013; Kwak and Lee, 1997; Blake and Carter, 2001). Furthermore, only one of the papers reviewed applied goal programming to healthcare in the African context (Agyei, Obeng-Denteh, and Andaam, 2015), and none to the South African context in particular.

There was only one paper reviewed which dealt with health resource allocation on a large scale the way this thesis aimed to do. Moon et al. (1989) successfully used a goal programming approach to allocate financial resources amongst Papua New Guinea's 19 provinces, and then went on to allocate the resources amongst seven health activities in each province. The idea was to develop a tool which could be used by policy makers to make allocation decisions more objectively and in line with equal access. This is a canonical example of the use of goal programming for equity in healthcare, which mirrors the motivation behind this thesis. However, there thus appears to be little literature pertaining to the allocating or shifting of tangible health resources rather than budgetary allocations on a country-wide scale. It is worth noting that this is not the case when one considers optimisation methods beyond goal programming. For example, Rauner and Bajmoczy (2003) developed a decision support system based on integer programming to allocate medical devices across different regions of Austria. Additionally, the field of humanitarian logistics presents numerous examples of optimisation used to allocate tangible health resources. For instance, Serrato-Garcia, Mora-Vargas, and Murillo (2016) developed a multi-objective optimisation model for resource allocation during disaster relief.

2.5 Simulation and optimisation in healthcare

As this thesis aims to combine disease modelling and goal programming to address an issue in public health, an examination of the historical use of simulation and optimisation together in the healthcare space is warranted. The vast majority of such applications entail discrete event simulation (DES) (Oddoye et al., 2009; Oliveira et al., 2020; Jerbi and Kamoun, 2009; Lal, Roh, and Huschka, 2016; Feng, Wu, and Chen, 2017; Lucidi et al., 2016; Chen et al., 2015; Ozcan, Tànfani, and Testi, 2017), and an iterative simulation-optimisation routine (Oliveira et al., 2020; Lucidi et al., 2016; Chen et al., 2015; Ozcan, Tànfani, and Testi, 2017). On the other hand, this thesis aimed to use simulation to generate key parameters for the optimisation problem, and to evaluate optimal solutions once they had been generated. Additionally, almost all sources reviewed looked at scheduling or resource allocation within a particular healthcare institute rather than on a multi-institute or national scale. Oliveira et al.

(2020) was the only exception: the study looked at outsourcing hospital beds from the private sector and reallocating public sector beds amongst Brazilian cities. However, their simulation assumption when demand exceeded capacity was that capacity would be temporarily stretched to cover the shortfall rather than dealing directly with the consequences of scarcity.

There were three papers found which combined simulation with goal programming in particular. Oddoye et al. (2009) used DES results to minimise metrics such as patient queue lengths in the medical assessment unit of a UK hospital. Jerbi and Kamoun (2009) also utilised DES in conjunction with goal programming to select optimal schedules in terms of, amongst other goals, minimising patients' waiting times in a Tunisian hospital. Meanwhile, Ahmed and Alkhamis (2009) combined Monte Carlo simulation with goal programming to reduce patients' overall time spent in an emergency department unit of a government hospital in Kuwait.

There is much precedent for the use of simulation and optimisation in the context of epidemic control. Malik and Sharomi (2017) conducted a review of such methods, noting goal programming as one of the key optimisation strategies. Long, Nohdurft, and Spinler (2018), using the 2014 Ebola outbreak as a case study, developed a multi-patch Susceptible-Infectious-Removed model with connectivity amongst regions which incorporated behavioural changes. The optimisation component compared four methods: a heuristic based on detected cases, a greedy R0-based policy, a myopic linear programme which worked in an iterative loop with the model, and a dynamic programming approach. The myopic linear programme performed the best across many scenarios and was very fast computationally. A key limitation of the study was that more complex models were not explored. Mbah and Gilligan (2011) similarly looked at emerging infectious disease outbreaks in multiple interconnected regions, this time using an Susceptible-Infectious-Removed-Susceptible model with two coupled sub-populations. They used the Pontryagin maximum principle to establish that the most efficient strategy was first to allocate resources to the more infected sub-population, and later to switch to allocating resources to the less infected subpopulation.

2.6 Conclusion

At the time of writing, COVID-19 has no cure and no vaccine approved for use. South Africa's population, largely reliant as it is on public health care, is subject to geographic variations in the quality of health care. This shows a clear need for resource sharing amongst South Africa's districts in the time of COVID-19. While many have used SEIR models to project the spread of COVID-19 in the African and South African contexts, few have addressed the idea of limited capacity, which is where this thesis contributed to the NCEM.

3

Data

There are crucial sets of data which support various aspects of this thesis. It is additionally worth examining the daily confirmed COVID-19 case data for South Africa, in order to understand the context more completely. This chapter will chronicle the collection and analysis of these data.

3.1 Collection

The four main datasets are outlined below:

1. The numbers of ICU and general hospital beds available in each district of SA were estimated and provided by the Health Economics and Epidemiology Research Office (HE2RO) (Van Den Heever, 2020).

2. The number of operational ambulances in each province was obtained in consultation with Nikhil Khanna and Tucker Bbosa from the Clinton Health Access Initiative (CHAI) (National Department of Health, 2020). The figure was missing for Gauteng, and was estimated by inflating the Western Cape's number according to the ratio of the two populations.

3. The cost of transporting patients via ambulance between districts was based on the National Reference Price List for Ambulance Services (RSA Government, 2008). The latest figures are from 2009; they were inflated by 40% in order to account for increases in the price of fuel. They were then decreased by 20% to reflect prices in the public sector, as advised by the experts at CHAI. The base fee for the first sixty minutes was assumed to apply to the first 100km of distance between two districts; thereafter, the per kilometre fee for long-distance travel was used.

4. The confirmed daily cases for SA were obtained from the GitHub repository maintained by the Data Science for Social Impact Research Group at the University of Pretoria (Data Science for Social Impact Research Group @ University of Pretoria, 2020).

3.2 Analysis

The numbers of ICU and general hospital beds are shown per 1000 of the population in Figure 3.1. It is immediately evident that the two were on very different scales;

the most resource-rich district in terms of ICU beds (Cape Town Metropolitan Municipality) had 0.15 beds per 1000 people, while in terms of general hospital beds, Johannesburg Metropolitan Municipality had 2.4 beds per 1000 people. Furthermore, there were twelve districts which did not have any ICU beds at all, while there were no districts which had no general hospital beds. This is intuitive, as ICU beds tend to be in short supply across Africa, and clearly highlights the need for mobility of either resources or patients amongst the districts in the context of a pandemic. It is also evident that resources tended to be concentrated in the densely populated urban hubs; that is, Johannesburg, Cape Town, and Durban, even when considering numbers adjusted for population. This further highlights the need to share resources with resource-poor districts.

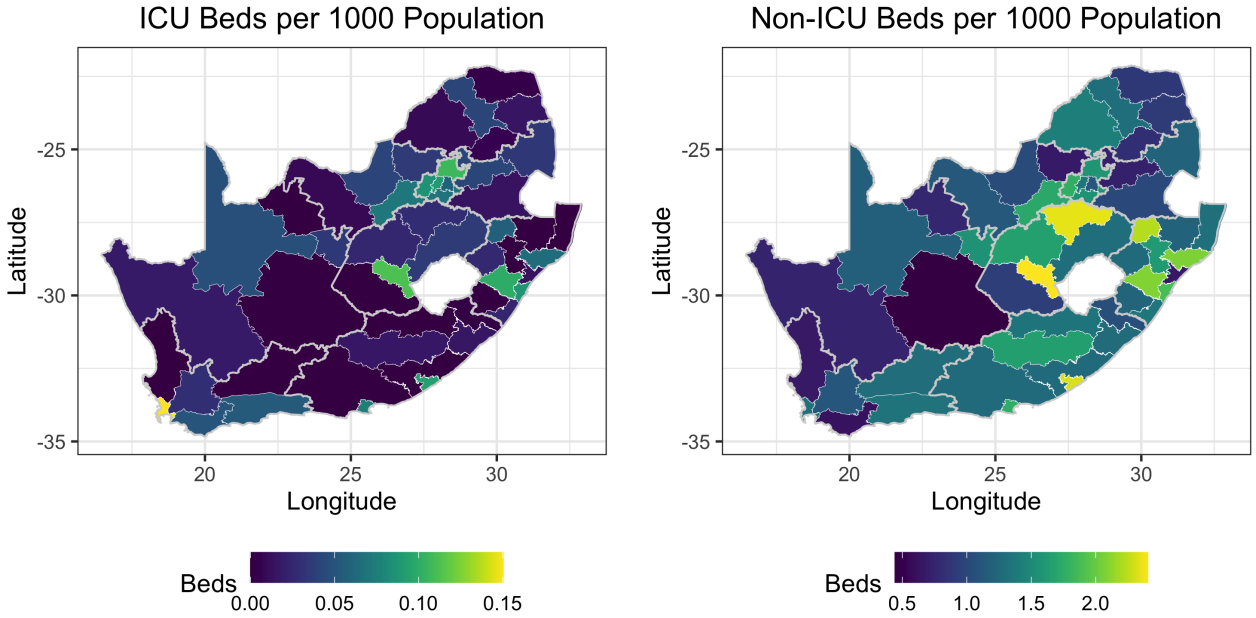

FIGURE 3.1: Numbers of ICU (left) and general hospital (right) beds in each district per 1000 people. The thick lines represent provincial borders, while thin lines separate districts.

Each province's number of operational ambulances is shown in Figure 3.2. Gauteng had the highest number of ambulances, followed by the Eastern Cape, KwaZulu-Natal, the Western Cape, and then the rest. This shows the same trend that provinces with large cities are the most resource-rich. However, in this case, the numbers of ambulances are absolute rather than per 1000 of the population. This is because it is not necessarily only population size which affects the need for ambulances, but also the geographical distances which the ambulances need to travel. It is worth noting, for example, that the largest province in terms of area - the Northern Cape - had one of the smallest counts of operational ambulances.

The estimated cost, in Rands, of transporting patients via ambulance between the districts intra-provincially is shown in Figure 3.4. This plot highlights the variation both in the number of districts per province and in the geographical distances involved in traversing each province. The Eastern Cape and Northern Cape stood out as having incredibly large distances amongst districts, causing the cost of a single trip

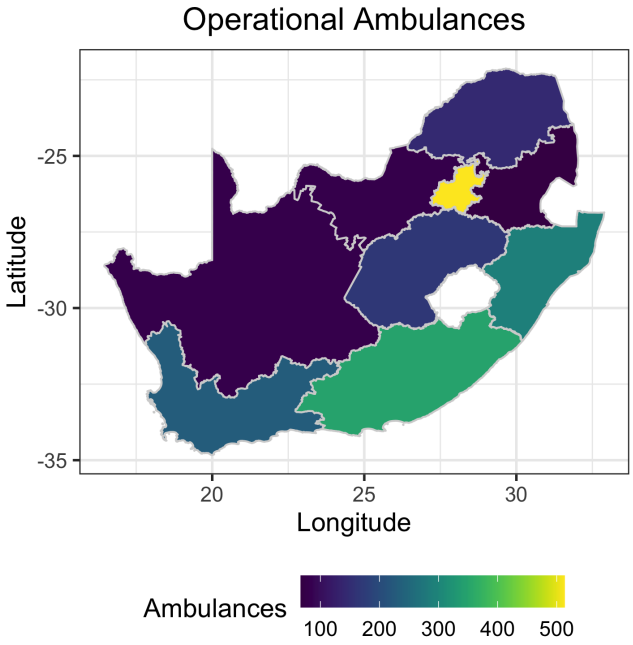

FIGURE 3.2: Number of operational ambulances in each province.

to exceed R12,500 in some cases. KwaZulu-Natal, on the other hand, has a great number of districts, but costs remained relatively low for the most part. Gauteng is the only province in which all distances fall within 100km, and thus only the initial fixed cost applies. This is likely a function of the fact that the distances are between centroids; in reality, there would be distances which would exceed 100km when road networks and extreme points are considered. The function used to calculate these costs is shown in Figure 3.3.

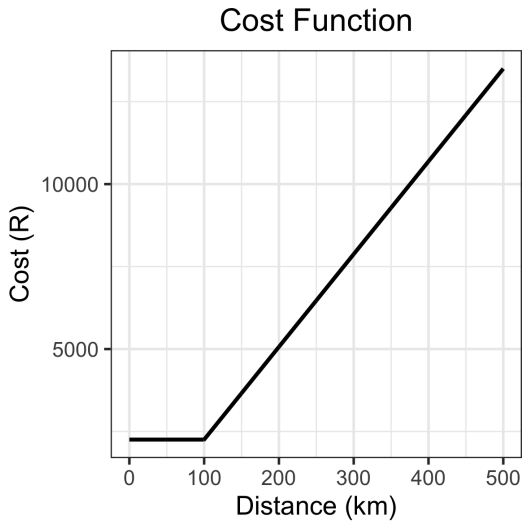

FIGURE 3.3: Cost function for transport of patients via ambulance.

Cost data were not available for the transportation of ICU beds from district to district; as such, an arbitrary scaled measure of distance (0.0001 of the kilometre distance) was used as a proxy for cost. Although this does not strictly qualify as a

dataset, the scaled distances between districts are shown in Appendix A. As expected, the scaled distances had the same visual trends as the costs of ambulance transport did. The most notable difference is that since there is no initial fixed distance, there was more variation in the proxy values for cost than there was in actual ambulance cost. This is immediately evident for provinces such as Gauteng and Mpumalanga. Note that while these proxy values are not actual costs, they are referred to as cost in this thesis for simplicity.

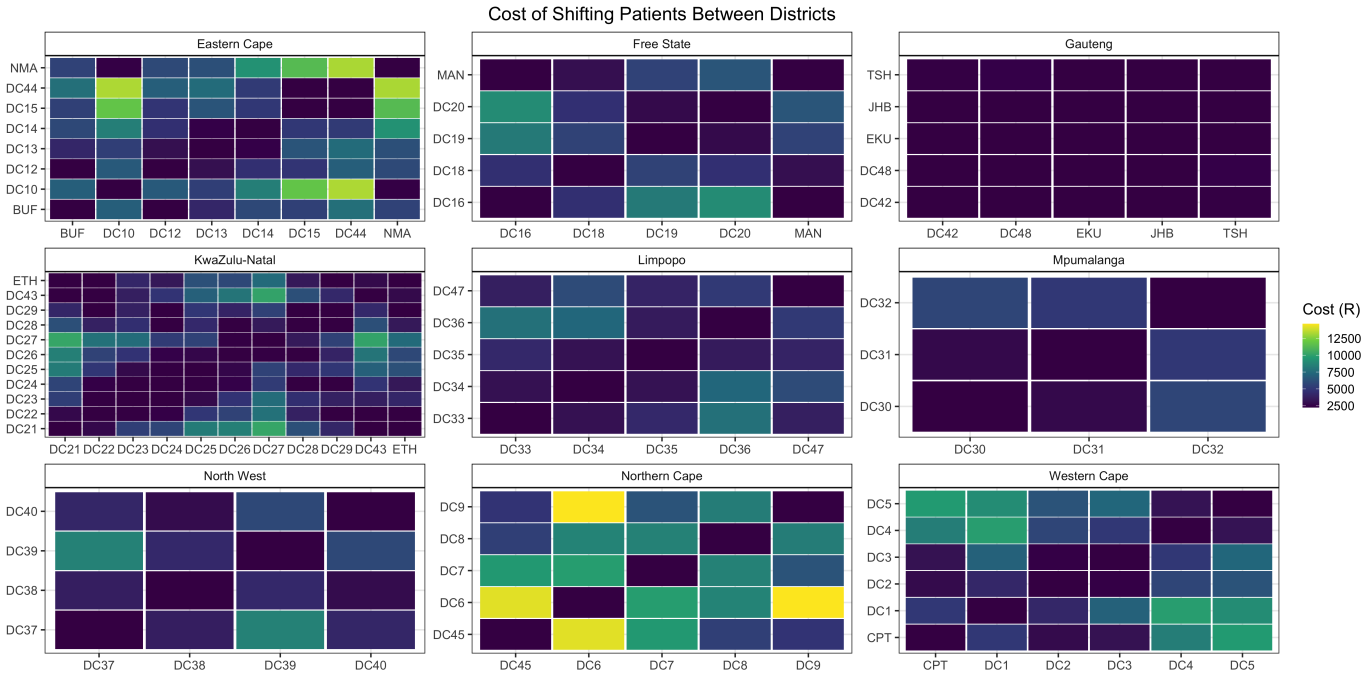

FIGURE 3.4: Cost (R) of transport via ambulance between districts within provinces.

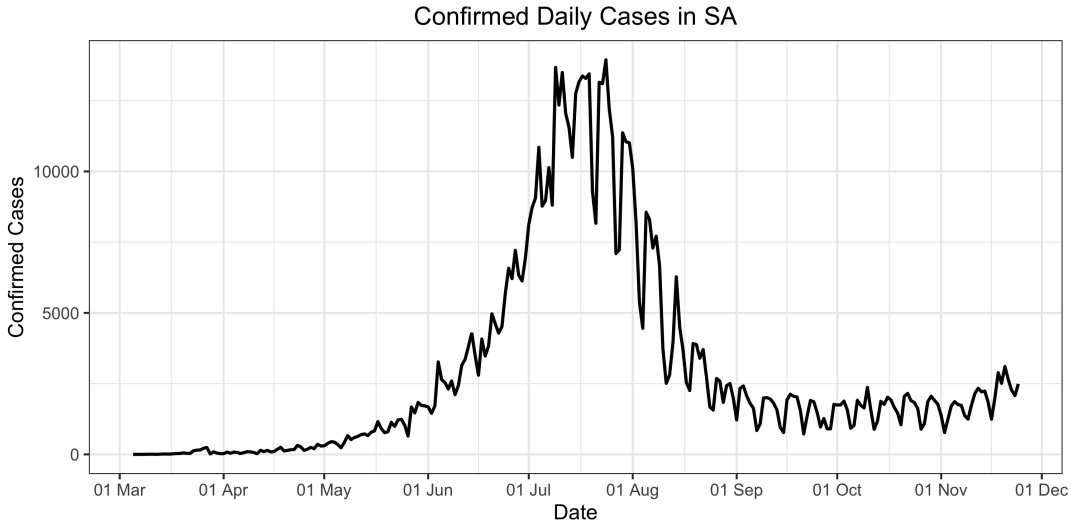

FIGURE 3.5: Number of daily confirmed cases in SA.

Finally, in setting the stage to proceed with this thesis, it is necessary to examine the COVID-19 pandemic as it was reported in South Africa, as of November 2020.

The daily confirmed cases are shown in Figure 3.5. At the peak of the first wave of COVID-19, the country reported approximately 12,000 new cases per day. The entire first wave began and ended between March and October. Although in reality, cases were beginning to increase again at the end of November, hinting towards a second wave, this thesis exclusively considered the first wave, and did not allow for the beginning of a second wave by the end of December.

4 Methodology

This chapter will expand on the topics of disease modelling and goal programming as they were applied to this thesis. Additionally, the framework interfacing disease modelling and goal programming will be introduced and motivated.

4.1 Disease modelling

This section will build the concept of an SEIR model and its components in a step-by-step fashion, at each point discussing this thesis's deviations from the NCEM, if any.

The SEIR model, first introduced by Kermack and McKendrick (1927), is a population-level compartmental model in which individuals transition amongst different compartments according to population-averaged rates and probabilities. Generalised SEIR models allow for any number of compartments and pathways, in addition to the proportions of susceptible (S), exposed (E), infectious (I), and removed (R). The transitions can be written as, and are solved as, a system of first order differential equations. The equations for a simple SEIR model are shown below, while the corresponding model diagram is depicted in Figure 4.1. The force of infection, or rate at which susceptible people become exposed to the virus, is represented by $\lambda = \beta I$, where β is known as the effective contact rate. The inverse of the incubation period is denoted by σ, while γ denotes the inverse of the recovery period.

$$\begin{aligned}
\frac{dS}{dt} &= -\beta IS \\
\frac{dE}{dt} &= \beta IS - \sigma E \\
\frac{dI}{dt} &= \sigma E - \gamma I \\
\frac{dR}{dt} &= \gamma I
\end{aligned}$$

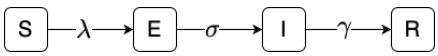

FIGURE 4.1: Simple SEIR model structure.

The force of infection is one of the most important and complex disease parameters in an SEIR model (Hens et al., 2010). It can be modified in order to take into account different factors affecting disease spread; this thesis deviated from the NCEM primarily via two adjustments to the force of infection. In simple terms, and without any adaptations, the force of infection can be calculated as the ratio of the number of infectious individuals in a particular district to the total number of (living) individuals in that district, multiplied by β, the effective contact rate. The effective contact rate is simply the number of contacts between people per unit time, which in this case is days, multiplied by the probability of transmission. These two components are not, however, identifiable. As such, the formula for the force of infection can be summarised as follows for the i^{th} district:

$$FOI_i = \beta_i I_i.$$

The NCEM is implemented over South Africa's 52 districts, and, in this thesis, movement amongst districts was incorporated via a simple gravity model, which scales the force of infection according to the inverse distance between the centroids of the districts (Barthelemy, 2010). This is a simplification of the NCEM, and implies that districts contribute to each other's levels of infection in a way that is inversely proportional to distance. The NCEM calculated movement according to a Vodacom dataset, to which the candidate was not able to gain access. Suppose D represents the 52 by 52 matrix of distances amongst the districts. Then the elements c_{ij} of a connectivity matrix C are calculated as follows:

$$c_{ij} = \frac{\frac{1}{1+hD_{ij}}}{\sum_j \frac{1}{1+hD_{ij}}}.$$

Thereafter, C is normalised so that the rows sum to one. The tuning parameter h controls the level of connectivity amongst the districts; the larger the distance between two districts, the less connected they are, and a larger h amplifies this effect, in turn slowing down spatial spread. The c_{ij} can be interpreted as probabilities: c_{ij} is the probability of moving from from district i to district j, of all possible destinations. The gravity model thus implicitly captures the effect of mobility without actually allowing populations to flow from one district to another. The benefit of this is that it avoids mixing populations into homogeneous compartments. Agent based models (ABMs) are better suited if one wishes to allow for explicit population movement.

In order to achieve comparable output to that of the NCEM, h was set to 0.01. Note that it was desirable to set tuning parameters such that results were comparable to the NCEM's, as this ensured that the model was producing realistic results. The formula for the force of infection is adjusted as follows:

$$FOI_i = \sum_j c_{ij}\beta_j I_j.$$

Thus, the rate at which susceptible people acquire the disease in district i is the sum of the forces of infection in all districts, weighted by the probability of moving to each district. Note that the diagonals of C are typically very close to one, indicating a high

probability that individuals would be found in their home district.

The second deviation from the NCEM was implemented by way of density-dependent scaling of the force of infection, which improves the way disease is assumed to spread by allowing it to spread faster in districts with higher population density (Hu, Nigmatulina, and Eckhoff, 2013). The formula for the force of infection worked with thus far is in fact based on a frequency-dependent interpretation of transmission, in which transmission is determined by the effective contact rate and the varying number of individuals in the population at any given time. Under a density-dependent interpretation, the denominator of the formula is instead set to the fixed population size, implying that transmission depends rather on the general density of a population. For example, a densely populated urban area would stay densely populated even while the current population fluctuates, as would a rural area remain sparsely populated. It is intuitive that disease would spread faster in the former context. However, as shown by Hu, Nigmatulina, and Eckhoff (2013), this effect is more pronounced when population sizes are small than it is when they are large. That is to say, the smaller the population in a particular district, the more likely it is that transmission will depend on density. Hu, Nigmatulina, and Eckhoff (2013) subsequently developed a hybrid approach which allows the effect of population density on transmission to saturate at a certain level. The force of infection for each district is thus multiplied by a density factor, as follows:

$$FOI_i = \sum_j c_{ij}\left(1 - \exp\left(-\frac{PD_j}{\rho}\right)\right)\beta_j I_j.$$

Population density in the j^{th} district is represented by PD_j, while ρ is a tuning parameter which controls the point at which the density-dependent part of transmission saturates and allows for frequency-dependent transmission to dominate. For the purposes of closely mirroring the output produced by the NCEM, but to produce a slightly lower attack rate (the proportion of the population which has been infected over the entire period of the model) which reflects the reduced effect of density for districts with larger populations, ρ was chosen to be twenty.

The structure of the generalised SEIR model, as designed by the SACMC, is shown in Figure 4.3. Individuals move from being susceptible to exposed, and are then either asymptomatic or are pre-symptomatic and develop symptoms. The majority of infections are mild and require no hospital treatment. For the purposes of this thesis, it was assumed that all of those who subsequently need treatment do seek treatment, which renders the $I_{S\bar{T}}$ compartment obsolete. This is because governments should plan around the potential need for services rather than the anticipated use thereof. Additionally, in the unprecedented situation of a global pandemic, it is impossible to predict what the use of services would be.

In the absence of capacity constraints, COVID-19 patients progress either to a general hospital ward or to ICU, where ICU is split according to ventilation needs and whether patients are destined to die or to be discharged. The primary contribution of this thesis is expressed in the addition of waiting chambers and accompanying capacity constraints. As seen in Figure 4.3, there is a chamber W_H for patients waiting for a general hospital bed, and W_V and $W_{\bar{V}}$ for patients waiting for ICU beds.

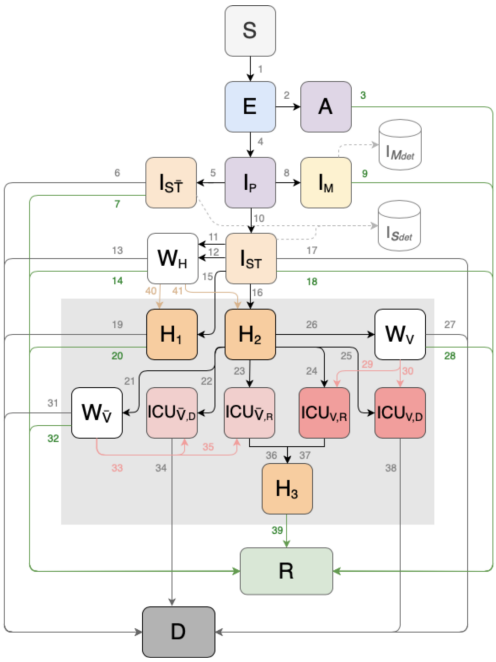

FIGURE 4.2: National COVID Epi Model structure (SA COVID-19 Modelling Consortium, 2020b). Note that the candidate's structural adjustments to the NCEM are included in this diagram as they were adopted and used by the SACMC.

When general hospital or ICU capacity is reached, patients are redirected to these waiting chambers. Patients either die, recover, or re-enter the treatment pathways after spending time in any one of the waiting chambers. In real terms, W_H represents patients who are not in the hospital system at all - they may be in a physical waiting room - while patients in W_V and $W_{\bar{V}}$ are assumed to be occupying general hospital beds while they wait. Finally, all individuals either die or join the removed state, in which they are no longer hospitalised nor infectious.

The flows to and from the waiting chambers for ICU are governed by the rates and probabilities shown in Table 4.2. The mechanism works in the same way for hospital admissions. The parameters are specified in Table 4.1. The probability of requiring ventilation is denoted by p_V, while τ_{ICU} is the inverse of the duration of progress to ICU, pd_{2V} and $pd_{2\bar{V}}$ are the probabilities of dying in ICU on and off ventilation respectively, and r_9 and r_{10} are the inverse of the times to recovery for those in ICU, on and off ventilation respectively. The scale factors sc_2 and sc_3 are used to inflate the probability of mortality - and in turn deflate the probability of recovery - to account for patients who are in need of treatment but cannot proceed owing to the non-availability of ICU beds. The $avail_2$ variable controls the mechanism of the capacity constraint in the way described by the Algorithm 1. The parameters μ_V and $\mu_{\bar{V}}$ represent the inverse of the times to death on and off ventilation respectively. The corresponding differential equations, for each district, are as follows:

$$\frac{\mathrm{dW_V}}{\mathrm{dt}} = (1-avail_2)p_V\tau_{ICU}H_2 - (1-min(sc_2pd_{2V},1))r_9W_V - min(sc_2pd_{2V},1)r_9W_V.$$

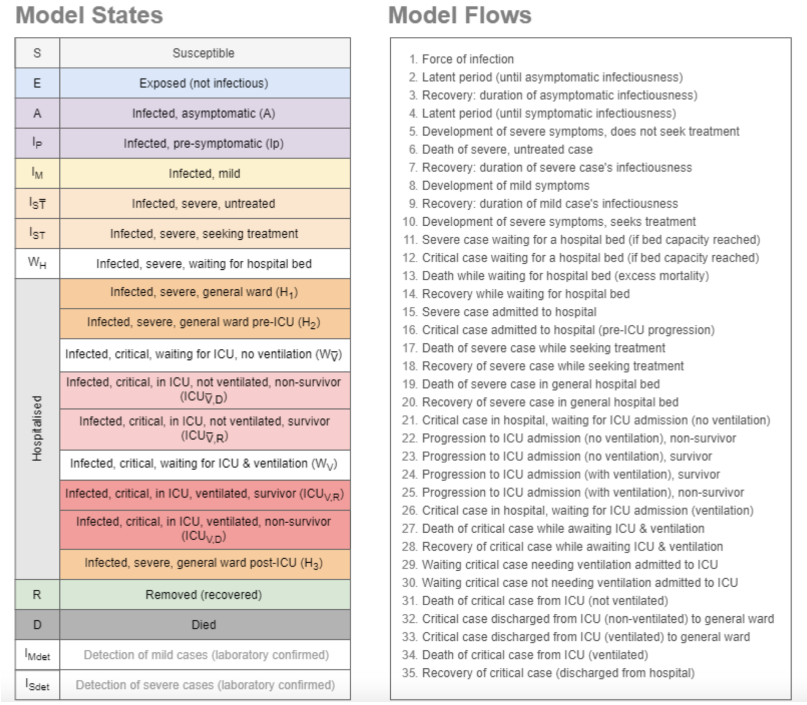

FIGURE 4.3: National COVID Epi Model legend of compartments and flows (SA COVID-19 Modelling Consortium, 2020b).

$$\frac{\mathrm{dW}_{\bar{V}}}{\mathrm{dt}} = (1-avail_2)(1-p_V)\tau_{ICU}H_2 - (1-min(sc_3pd_{2\bar{V}},1))r_{10}W_{\bar{V}} - min(sc_3pd_{2\bar{V}},1)r_{10}W_{\bar{V}}.$$

Algorithm 1: Capacity constraint algorithm

for *each district* **do**
  Compute current occupied ICU, $ICU_{occupied} = ICU_{\bar{V},D} + ICU_{\bar{V},R} + ICU_{V,D} + ICU_{V,R}$;
  Compute patients proposed to enter ICU, proposed_enter $= \tau_{ICU} * H_2$;
  Compute patients proposed to exit ICU, proposed_exit $= \mu_V * ICU_{V,D} + r_9 * ICU_{VR} + \mu_{\bar{V}} * ICU_{\bar{V},D} + r_{10} * ICU_{\bar{V},R}$;
  Set boolean violation = ICU_occupied - proposed_exit + proposed_enter > icu_capacity;
  if *violation is TRUE* **then**
    avail = (icu_capacity - ICU_occupied + proposed_exit) / proposed_enter;
  else
    avail = 1;
  end
end

One of the most important measures in the realm of infectious diseases is the basic reproductive number, known as $\mathcal{R}_0$. This quantity represents *the average number of secondary cases produced by one infected individual introduced into a population of susceptible individuals* (Driessche, 2017). It can be calculated as the product of the probability of transmission given a contact between an infectious individual and a susceptible individual, the average number of contacts between susceptible and infectious individuals, and the duration of infectiousness. Intuitively, when $\mathcal{R}_0 > 1$, infection is increasing. It can be shown that the below equation holds for the basic

Parameter	Default Value	Lower Parameter	Upper Parameter
p_V	$*0.402$	0.8	1.2
τ_{ICU}	365	183	730
pd_{2V}	$*0.548$	0.8	1.2
$pd_{2\bar{V}}$	$*0.303$	0.8	1.2
r_9	19	15	46
r_{10}	73	41	183
sc_2	2		
sc_3	1		
μ_V	26	20	52
$mu_{\bar{V}}$	33	30	73

TABLE 4.1: Specification of the parameters involved in the ICU capacity constraint flows. All parameters were sampled using a triangular distribution for the purpose of stochasticity (see the next sub-section), except for sc_2 and sc_3, which remained fixed. *These parameters differed by province, and the Default Value listed is the mean. For these parameters, each default value was multiplied by a draw of the triangular distribution shown, and forced to be no greater than one.

Flow	Transition Value	Flow	Transition Value
$H_2 \rightarrow W_V$	$(1 - avail_2) * p_V * \tau_{icu}$	$H_2 \rightarrow W_{\bar{V}}$	$(1 - avail_2) * (1 - p_V) * \tau_{ICU}$
$W_V \rightarrow D$	$min(sc_2 * pd_{2V}, 1) * r_9$	$W_{\bar{V}} \rightarrow D$	$min(sc_3 * pd_{2\bar{V}}, 1) * r_{10}$
$W_V \rightarrow R$	$(1 - min(sc_2 * pd_{2V}, 1)) * r_9$	$W_{\bar{V}} \rightarrow R$	$(1 - min(sc_3 * pd_{2\bar{V}}, 1)) * r_{10}$

TABLE 4.2: Flows and transition values pertaining to the ICU capacity constraint.

SEIR model indicated in Figure 4.1. This is in line with the definition of $\mathcal{R}_0$, as the effective contact rate is the number of contacts between people per unit time, multiplied by the probability of transmission, and $\frac{1}{\gamma}$ represents the duration of the recovery period, or the time it takes to leave the infectious compartment.

$$\mathcal{R}_0 = \frac{\beta}{\gamma}$$

$\mathcal{R}_0$ can be calculated for the NCEM, for a particular district. The derivation begins with stating the equations relating to exposure and infection as follows, where I^* represents the infectious reservoir; that is, $I^* = (\zeta A + \zeta I_P + I_M + I_{ST} + W_H + I_{S\bar{T}})$ (note that the simplifying assumption is made that once hospitalised, a person is no longer infectious):

$$
\begin{aligned}
\frac{dE}{dt} &= \beta I^* S - \gamma_1 E \\
\frac{dA}{dt} &= p_A \gamma_1 E - r_1 A \\
\frac{dI_P}{dt} &= (1 - p_A)\gamma_1 E - \gamma_2 I_P \\
\frac{dI_M}{dt} &= p_M \gamma_2 I_P - r_2 I_M \\
\frac{dI_{ST}}{dt} &= (1 - p_M) p_{TS} \gamma_2 I_P - avail_1 \tau_S I_{ST} \\
\frac{dW_H}{dt} &= (1 - avail_1)\tau_S I_{ST} - r_8 W_H \\
\frac{dI_{S\bar{T}}}{dt} &= (1 - p_M)(1 - p_{TS})\gamma_2 I_P - r_8 I_{S\bar{T}}
\end{aligned}
$$

Note that the force of infection is written as βI^* for simplicity, and that ζ is used to scale asymptomatic or pre-symptomatic compartments to be less infectious than fully symptomatic compartments. These equations are set to zero to represent equilibrium - in which no compartment is growing or shrinking - and then solved such that each compartment can be written in terms of I_P:

$$
\begin{aligned}
E &= \frac{\gamma_2}{(1 - p_A)\gamma_1} I_P \\
A &= \frac{p_A \gamma_1}{r_1} E = \frac{p_A \gamma_1}{r_1} \frac{\gamma_2}{(1 - p_A)\gamma_1} I_P \\
I_{ST} &= \frac{(1 - p_M) p_{TS} \gamma_2}{avail_1 \tau_S} I_P \\
W_H &= \frac{(1 - avail_1)\tau_S}{r_8} I_{ST} = \frac{(1 - avail_1)\tau_S}{r_8} \frac{(1 - p_M) p_{TS} \gamma_2}{avail_1 \tau_S} I_P \\
I_{S\bar{T}} &= \frac{(1 - p_M)(1 - p_{TS})\gamma_2}{r_8} I_P
\end{aligned}
$$

In order for infection to be increasing, it is required that $\frac{dE}{dt} > 0 \iff I^* > \frac{\gamma_1}{\beta} E$. Then:

$$
\begin{aligned}
(\zeta A + \zeta I_P + I_M + I_{ST} + W_H + I_{S\bar{T}}) &> \frac{\gamma_1}{\beta} E \\
(\zeta A + \zeta I_P + I_M + I_{ST} + W_H + I_{S\bar{T}}) &> \frac{\gamma_1}{\beta} \frac{\gamma_2}{(1-p_A)\gamma_1} I_P \\
(\zeta A + \zeta I_P + I_M + I_{ST} + W_H + I_{S\bar{T}}) &> \frac{\gamma_2}{\beta(1-p_A)} I_P \\
\beta\Bigg(\frac{p_A\gamma_1}{r_1} I_P + I_P + \frac{p_M\gamma_2}{r_2} I_P + \frac{(1-p_M)p_{TS}\gamma_2}{avail_1\tau_S} I_P + \\
\frac{(1-avail_1)\tau_S}{r_8} \frac{(1-p_M)p_{TS}\gamma_2}{avail_1\tau_S} I_P + \frac{(1-p_M)(1-p_{TS})\gamma_2}{r_8} I_P \Bigg) &> \frac{\gamma_2}{(1-p_A)} I_P \\
\beta\Bigg(\frac{\zeta p_A}{r_1} + \frac{\zeta(1-p_A)}{\gamma_2} + \frac{(1-p_A)p_M}{r_2} \\
+ \frac{(1-p_A)(1-p_M)p_{TS}}{\tau_S} + \frac{(1-p_A)(1-p_M)(1-p_{TS})}{r_8} \Bigg) &> 1
\end{aligned}
$$

The last step holds because in a population that is fully susceptible save for one infection, capacity has not been breached and thus $avail_1 = 1$. Hence, for infection to be increasing, the left hand side quantity needs to be greater than one, which implies that this quantity is in fact $\mathcal{R}_0$. Therefore:

$$
\mathcal{R}_0 = \beta\Bigg(\frac{\zeta p_A}{r_1} + \frac{\zeta(1-p_A)}{\gamma_2} + \frac{(1-p_A)p_M}{r_2} + \frac{(1-p_A)(1-p_M)p_{TS}}{\tau_S} + \frac{(1-p_A)(1-p_M)(1-p_{TS})}{r_8} \Bigg).
$$

This result can be understood in the context of the definition of $\mathcal{R}_0$, as was done for the simple SEIR model. The effective contact rate represents, once more, the product of the number of contacts between people per unit time and the probability of transmission. Ignoring numerators, each fraction in the parentheses equates to the duration of infectiousness for the respective compartments A, I_P, I_M, I_{ST}, (W_H), and $I_{S\bar{T}}$. The numerators dictate the path which must be followed, in terms of probabilities, to reach each of those compartments. For example, for an individual to be an asymptomatic person in A, the probability of being asymptomatic, p_A, must be applied. An individual who is pre-symptomatic in I_P requires the converse i.e. the probability of not being asymptomatic, $(1 - p_A)$. An individual who has a mild infection in I_M needs to be not asymptomatic and mild, hence the application of both $(1 - p_A)$ and p_M. The rest of the compartments can be rationalised in the same way. Note that the fraction pertaining to W_H disappeared when the substitution $avail_1 = 1$ was made. This makes sense, because in a completely susceptible population, there is no notion of capacity's being reached and thus even the single infected person could not be in a waiting chamber. Finally, the inclusion of ζ has the function of reducing transmission probability for the asymptomatic and pre-symptomatic compartments. For all districts, it can be shown that $\mathcal{R}_0 \approx 2.5$.

In this thesis, the model was implemented in `R`, and the differential equation solver `deSolve` was employed, using the Euler method (Soetaert, Petzoldt, and Setzer, 2010). All parameters were set as per the SACMC's code (SA COVID-19 Modelling Consortium, 2020a).

4.2 Goal programming

In the realm of multi-objective decision making, goal programming is an approach in which the decision maker specifies the desired level of performance on each objective, known as the goal. The aim is to get each objective as close as possible to its respective goal value, and the difference between the two is known as the deviation, say d_i for the deviation from the goal on objective i. The deviations are weighted according to weights w_i, which reflect the relative importance of a one-unit deviation from the stated goal on each objective. These weights therefore capture both scaling differences and differences in the importance of good performance on each objective. Some combination of the weighted deviations is then minimised.

Chebyshev goal programming involves minimising the largest weighted deviation, denoted by Δ. An improved approach is one in which $\Delta + \epsilon \sum_{i=1}^{Q} w_i d_i$ is minimised, where ϵ is some small number and Q is the number of goals (Stewart, 2007). The improved approach helps to differentiate between two solutions in which the maximum deviations are the same. Suppose, for example, there are two solutions with a maximum deviation of 100. Further suppose that in the first solution, the deviation from the rest of the objectives is zero, but in the second solution, this deviation is 90. Clearly, the first solution is better; however, the original Chebyshev formulation would not differentiate between the two solutions. The improved Chebyshev approach additionally minimises the weighted sum of all deviations, which allows it to choose the better solution in this case.

This thesis used the improved Chebyshev approach to minimise the largest shortfall in beds across districts, by deciding on a monthly schedule of shifts of ICU beds from April to December. Each month, the disease model was used to look forward and assess ICU bed needs for the next month, which were then fed into the goal programming problem as goals. The goal programming problem then produced an optimal solution of shifting ICU beds between districts. While the disease modelling only allows for the assessment of ICU beds, from a goal programming perspective it is easy to translate this to shifting patients or healthcare workers, because both are in some fixed ratio to beds. Define the following, for $P = 52$ districts:

- $a_i :=$ the available units of the resource in the i^{th} district, $\forall\ i = 1, ..., P$
- $g_i :=$ the goal units of the resource in the i^{th} district, $\forall\ i = 1, ..., P$
- $c_{ij} :=$ the unit cost of shifting the resource between districts i and j, $\forall\ i = 1, ..., P$ and $j = 1, ..., P$
- $n_{ij} = 1$ if districts i and j are in the same province and $n_{ij} = 0$ otherwise $\forall\ i, j$
- $m :=$ the minimum shift allowed (here $m = 5$, which was an arbitrary choice for the running of this problem, but allows for alteration in future works)
- $M :=$ the maximum shift allowed (here $M = 1000$)
- $I_1, ..., I_9 :=$ the provincial sets of districts, in alphabetical order of provinces
- $A_1, ..., A_9 :=$ the number of operational ambulances in each province
- $\delta_{ij} :=$ binary indicators denoting whether resources are to be shifted from district i to district j, $\forall\ i = 1, ..., P$ and $j = 1, ..., P$

- x_{ij} as the units of the resource to be shifted from district i to district j, $\forall\ i = 1, ..., P$ and $j = 1, ..., P$

Then the goal programming problem is to minimise the following:

$$\Delta\ +\ \epsilon\Big\{\sum_{i=1}^{P} w_i d_i + w_C\gamma\Big\}$$

The deviations from the beds-related goals are the d_i, while γ represents the deviation from the cost minimisation goal. Assuming equal weightings for all beds-related goals, this can be simplified to:

$$\Delta\ +\ \epsilon\Big\{w_B\sum_{i=1}^{P} d_i + w_C\gamma\Big\}$$

This assumption ensures that all districts are treated the same, in that the loss of a life in one district is no worse than it is in another. Note that there is no index for time in the formulation of the problem, because each month's goal programming problem is solved independently.

The problem is subject to the constraints summarised in Table 4.3. While the optimisation framework is based on the numbers of ICU beds available and shifted, this can be translated to patients considering that there is a one-to-one ratio between patients and ICU beds. The only difference in the framework is the inclusion or exclusion of constraint (8) in 4.3. These constraints can be understood as follows:

1. In order for Δ to represent the maximum weighted deviation, it must be greater than or equal to all weighted shortage-related deviations.
2. In order for Δ to represent the maximum weighted deviation, it must also be greater than or equal to the cost deviation.
3. The deviations d_i can be understood as the shortfall once the available units of the resource in district i are updated according to shifts into and out of that district. This shortfall is bad if it is positive and good if it is negative, hence it should be minimised.
4. The deviation γ represents the surplus in total cost above zero. This surplus is good if it is negative and bad if it is positive, hence it should be minimised.
5. All shifts should be between the minimum and maximum allowable numbers.
6. The resources shifted from district i should be less than or equal to the available resources in that district.
7. The binary indicator decision variables allow for resources to be shifted only between different districts.
8. The resources shifted from all districts in the k^{th} province should be less than or equal to the number of operational ambulances in the k^{th} province. This constraint is only applied when the intention is to shift patients as opposed to ICU beds.

9. Resources should only be shifted between districts in the same province. This constraint can easily be excluded to facilitate a scenario in which resources can be shifted nationally.

Constraints
1. $\Delta \geq w_i d_i \ \forall \ i$
2. $\Delta \geq w_C \gamma$
3. $a_i - \sum_j x_{ij} + \sum_j x_{ji} + d_i \geq g_i \ \forall \ i$
4. $\sum_i \sum_j c_{ij} x_{ij} - \gamma = 0$
5. $m\delta_{ij} \leq x_{ij} \leq M\delta_{ij} \ \forall \ i, j$
6. $\sum_j x_{ij} \leq a_i \ \forall \ i$
7. $\delta_{ij} = \{0, 1\} \ \forall \ i \neq j$ and $\delta_{ij} = 0 \ \forall \ i = j$
8. $\sum_{i \in I_k} \sum_j x_{ij} \leq A_k \ \forall \ k = 1, ..., 9$
9. $\delta_{ij} \leq n_{ij} \ \forall \ i, j$

TABLE 4.3: Constraints of goal programming problem.

The weights w_B and w_C are pieces of preference information which would usually be provided by the decision maker. In this case, the weights define the units of (somewhat arbitrary) cost the decision maker is willing to give up for a bed in the case of shifting beds, and the amount of money (in Rands) that the decision maker is willing to give up for a bed in the case of shifting patients. In the absence of access to decision makers, three trade-offs were chosen via a trial-and-error process of exploring different w_C, the results of which are shown in Figure 4.4. Note that the stochastic nature of these plots is explained in the next sub-section. The idea was to choose three weighting configurations for shifting beds, and three for shifting patients. One out of the three configurations should be biased towards minimising cost, and another should be biased towards minimising bed shortages. The third weighting configuration should be somewhere in between, and is referred to as the 'intermediate' configuration.

The case of shifting beds was tested with weights at regular intervals from zero to one, which is equivalent to starting with zero weight placed on cost, and ending with zero weight placed on bed shortages. When cost was weighted between zero and 0.4, the results were virtually the same in that deaths were very low. As this weight was increased, focus was shifted enough to the cost goals such that shifts were reduced and deaths thereby increased. Intuitively, when cost was given full weight, the results directly mirrored those of the baseline scenario before any optimisation. Finally, this progression from minimal deaths to the baseline scenario is visible on quite a coarse weighting scale. In contrast, when patients were shifted, the scale that was necessary to produce a similar progression was relatively fine. Similarly to the case of beds, the first four cost weightings produced comparable results, and only at the fifth weighting did deaths begin to increase. The weightings 1×10^{-5}, 1×10^{-4}, and 1×10^{-3} were chosen for shifting patients, while for shifting beds the final choices were 0.4, 0.6, and 0.8.

As mentioned before, the significance of the weights lies in the trade-offs that they imply. Suppose that, in the case of shifting patients, $w_C = 1 \times 10^{-5} = \frac{1}{100000}$ and $w_B = \frac{99999}{100000}$. Then $\Delta = \max\{\frac{99999}{100000} \sum_i d_i, \frac{1}{100000}\gamma\}$. If $\sum_i d_i = 1$ and $\gamma = 0$, or if $\sum_i d_i = 0$ and $\gamma = 99999$, $\Delta = \frac{99999}{100000}$. Hence a R99,999 deviation in cost is equivalent to a one unit deviation in beds, or the decision maker would be willing to give up R99,999 for a bed. It can similarly be shown that $w_C = 1 \times 10^{-4}$ implies that a bed

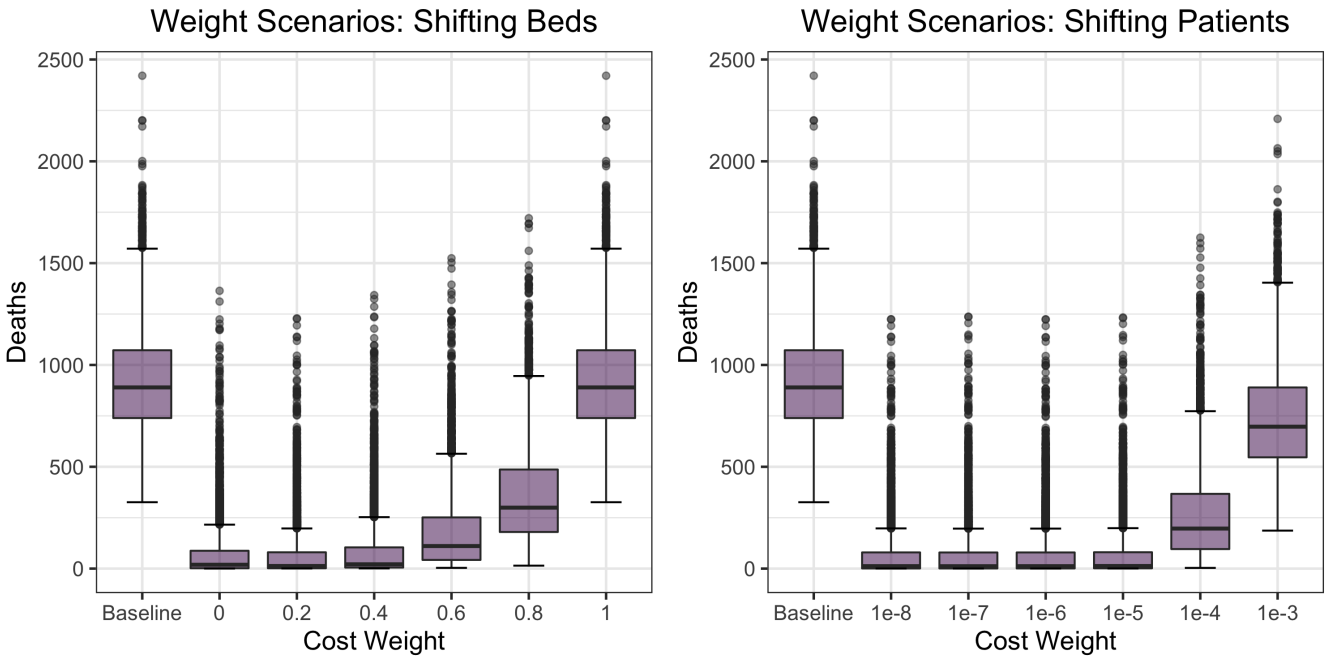

FIGURE 4.4: Experimentation with goal programming weights for shifting beds (left) and patients (right).

is worth R9,999, and that $w_C = 1 \times 10^{-3}$ implies that a bed is worth R999. These trade-offs are less intuitive in the case of shifting beds, where the cost is in arbitrary units. However, it can be shown that $w_C = 0.4$ implies that a three unit deviation in cost is equivalent to a two unit deviation in beds, or that the decision maker would be willing to give up 1.5 units of cost for a bed. It can similarly be shown that $w_C = 0.6$ and $w_C = 0.8$ imply that a two unit deviation in cost is equivalent to a three unit deviation in beds, and a one unit deviation in cost is equivalent to a four unit deviation in beds, respectively.

In this thesis, goal programming was implemented using the `Rglpk` package (Theussl and Hornik, 2019). The `R` implementation of this package does not allow for adjustment of tolerance or of stopping conditions; as such, a time limit on solving was set to five minutes. This means that if an optimal solution was not found within five minutes, the search would be halted and the best solution at that time would be used. This was a practical decision in the interests of the time constraints imposed on this thesis.

4.3 Combining disease modelling and goal programming

The framework to interface the disease modelling and optimisation components works as shown in Figure 4.5. There are two main custom aspects to the method. Firstly, the percentile of maximum ICU beds used for goals can be adjusted depending on how conservative one wishes to be. Note that the goals are intended to reflect demand for ICU beds; however, this demand is uncertain and risk attitudes must be taken into account. This method allows the decision maker to choose percentiles of the demand distribution from the simulation model. Secondly, weights can be adjusted to change

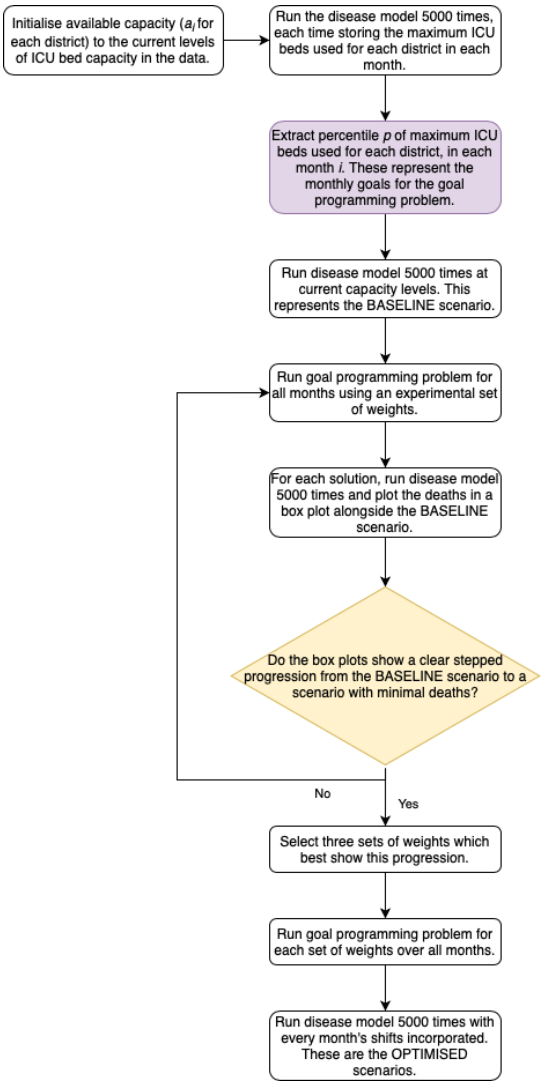

FIGURE 4.5: Disease modelling and goal programming framework.

the relative importance of the cost and beds goals.

Note that the goals are derived from running the disease model for the entire period (March to December 2020), at the beginning of the period. Although one might presume that in reality, the model would only be used to project a month at a time, and real-world data would be incorporated as the epidemic unfolds, this was in fact an infeasible notion in the South African context. Updated hospital admissions data were not available in real time or even with a small enough lag to be useful; it is only at the time of writing, after the first wave of the epidemic has completely passed, that such data are fully available.

4.4 Scenarios and sensitivity

A total of twelve scenarios were designed for this thesis. Six of them pertain to the monthly shifting of ICU beds intra-provincially under various weighting and goal scenarios, as shown in Table 4.4. Identical scenarios for the case of shifting patients via

ambulances are shown in Table 4.5. Monthly schedules of shifts in both cases work as shown in Figure 4.5.

When the diagram in Figure 4.5 indicates that the disease model should be run 5000 times, this means that the model should be run stochastically. That is, for each run, the disease parameters are sampled from appropriate distributions, as per the SACMC's code (SA COVID-19 Modelling Consortium, 2020a). Although the runs are stochastic, the same set of 5000 random seeds were used for each scenario, to ensure each set of 5000 runs corresponded to the same sampled parameters for each run.

Table 4.4: Scenario codes for monthly provincial shifting of beds.

	Weightings		
Goals Percentile	Intermediate	Cost-Focused	Beds-Focused
95^{th}	95-I-B	95-C-B	95-B-B
50^{th}	50-I-B	50-C-B	50-B-B

Table 4.5: Scenario codes for monthly provincial shifting of patients.

	Weightings		
Goals Percentile	Intermediate	Cost-Focused	Beds-Focused
95^{th}	95-I-P	95-C-P	95-B-P
50^{th}	50-I-P	50-C-P	50-B-P

The design of this framework and its components lends itself to sensitivity analysis in two ways. Firstly, the running of the simulation model stochastically allows for uncertainty to be captured. Secondly, sensitivity analysis was implicitly conducted through the different scenarios, in that the sensitivity to the goals percentile and the relative weightings are built into the scenarios. A final note related to sensitivity analysis is the issue of validation, and it is worth mentioning that the output of the simulation model was continually checked against that of the NCEM, as a way of validating the results.

5

Results

With the methodology established, the next step is to examine the results of both the disease modelling and the optimisation components thereof. This chapter will first explore various outcomes related to the disease model, and then the results of the disease modelling and goal programming framework.

5.1 Disease modelling

In this section, the following key findings will be shown:

- The NCEM projected the first wave to be finished by December 2020.
- General hospital and ICU bed capacities were projected not to be breached, and to be breached in two provinces, respectively.
- KwaZulu-Natal exemplified the need for intra-provincial resource sharing.
- Deaths due to the non-availability of ICU beds constituted as much as 60% of total mortality.
- Deaths due to the non-availability of ICU beds decreased monotonically with increasing number of ICU beds.

5.1.1 The NCEM projected the first wave to be finished by December 2020

At the time that the analysis for this thesis was being conducted, the projections of the NCEM were as shown in Figure 5.1. It is clear that the first wave was projected to be finished by December 2020. Although in reality, the second wave would have peaked within the window of these plots, this was due to the emergence of the lineage B.1.351, which was as yet unknown and thus not modelled by the NCEM at this point.

5.1.2 General hospital and ICU bed capacities were projected not to be breached, and to be breached in two provinces, respectively

The first portion of results to examine is that of the capacity mechanism in the NCEM. Figures 5.2 and 5.3 demonstrate, aggregated at a provincial level, the general and ICU beds in use with and without limited capacity. Limited capacity is when patients are refused a bed once capacity is reached, depicting actual use of beds, while unlimited

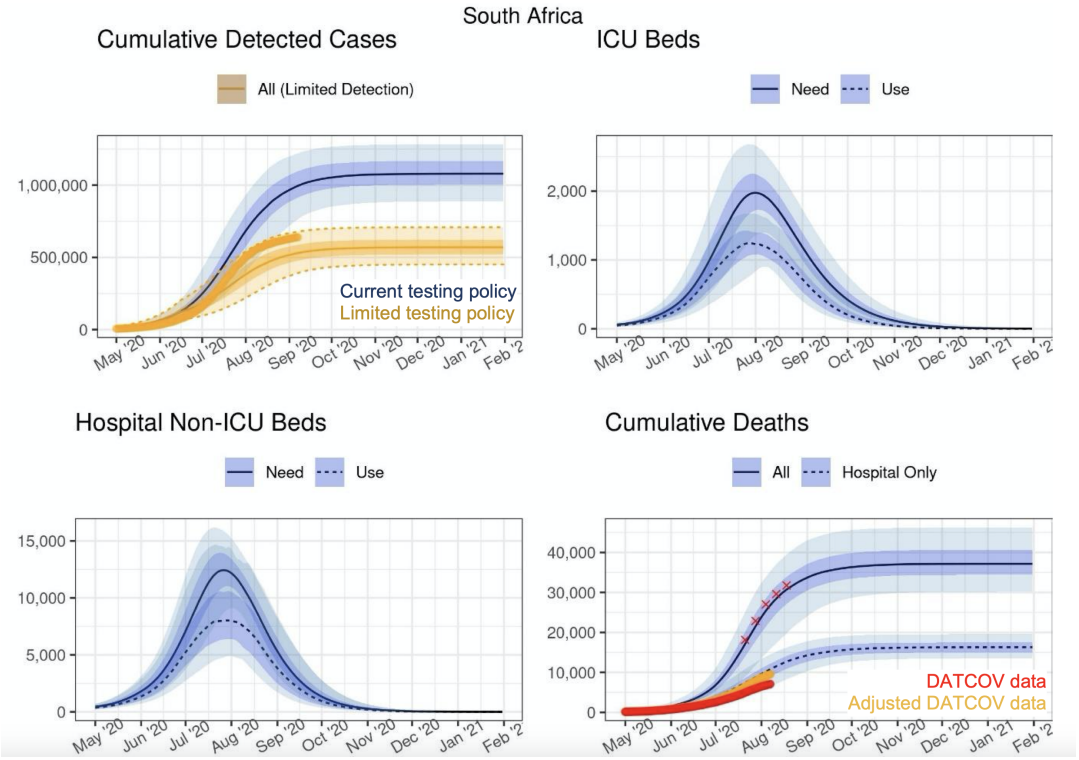

FIGURE 5.1: National projections of the NCEM (SA COVID-19 Modelling Consortium, 2020b). The top-left plot shows projected cumulative detected cases under the testing policy at the time (blue), and a policy of only detecting hospitalised cases as of mid-June (orange). The top-right and bottom-left plots show the projected need and use of ICU and general hospital beds respectively. Need refers to the full demand for beds, while use refers to the actual number used given limited capacity. The bottom-right plot shows projected cumulative deaths against available data. More information regarding these plots is available in the SACMC's September report (SA COVID-19 Modelling Consortium, 2020b).

capacity depicts the need for beds assuming no limit. The beds in use measures are expressed as percentages of total capacity in each province. Clearly, the non-ICU beds in use measure did not exceed capacity in any of the provinces; hence, the curves for limited and unlimited capacity overlapped completely. Additionally, non-ICU bed usage did not exceed fifty percent in any instance. This does not imply that there were no individual districts which ran out of capacity, but if so, there were unlikely to be drastic overflows. Bed utilisation in percentage terms was highest in Mpumalanga and lowest in the Northern Cape. The case of ICU beds in use was different. The Free State and KwaZulu-Natal both exceeded capacity substantially, while the Eastern Cape and Mpumalanga were close to doing so, with capacity utilisation at more than ninety percent. It can also be observed that in several provinces, even though capacity was not exceeded, the beds in use in the limited and unlimited cases did not overlap entirely. This suggests that there were districts for which capacity was reached.

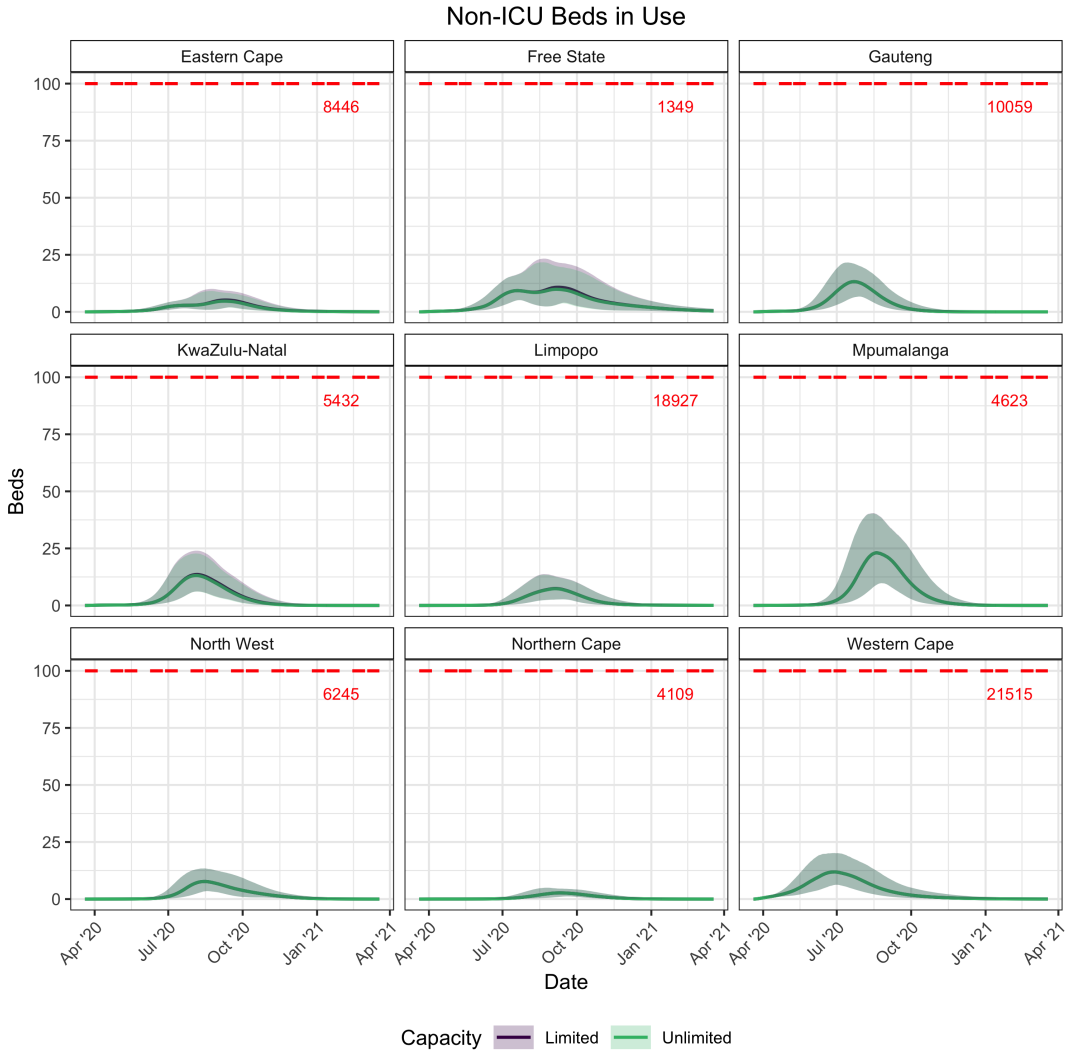

FIGURE 5.2: Non-ICU beds in use over time per province. The red dashed lines represent capacity. The bands represent approximate 95% confidence intervals.

5.1.3 KwaZulu-Natal exemplified the need for intra-provincial resource sharing

Developing a more nuanced understanding of the district-level dynamics at play necessitates a closer look at one of the provinces. Figure 5.4 demonstrates the ICU beds in use in KwaZulu-Natal, by district, excluding the districts in which there is zero capacity. This is because, again, the beds in use measure is expressed as a percentage of total capacity in each district. Five of these districts ran out of ICU beds, while the other two exhibited utilisation levels in excess of eighty percent. While several districts ran out of capacity, it is further worth noting that the extent to which districts exceeded capacity varied significantly. In DC25 and DC28 - as defined in List 1 - the bed utilisation was well below two hundred percent. In DC29, however, the utilisation reached a staggering six hundred percent. This disparity highlights the need for intra-provincial resource sharing. Note that this plot is available for the rest of the provinces in Appendix B. These plots indicate that ICU capacity was breached in at least one district in the Eastern Cape, Limpopo, and Mpumalanga, while in the Northern Cape and the Western Cape, at least one district did not have any ICU beds.

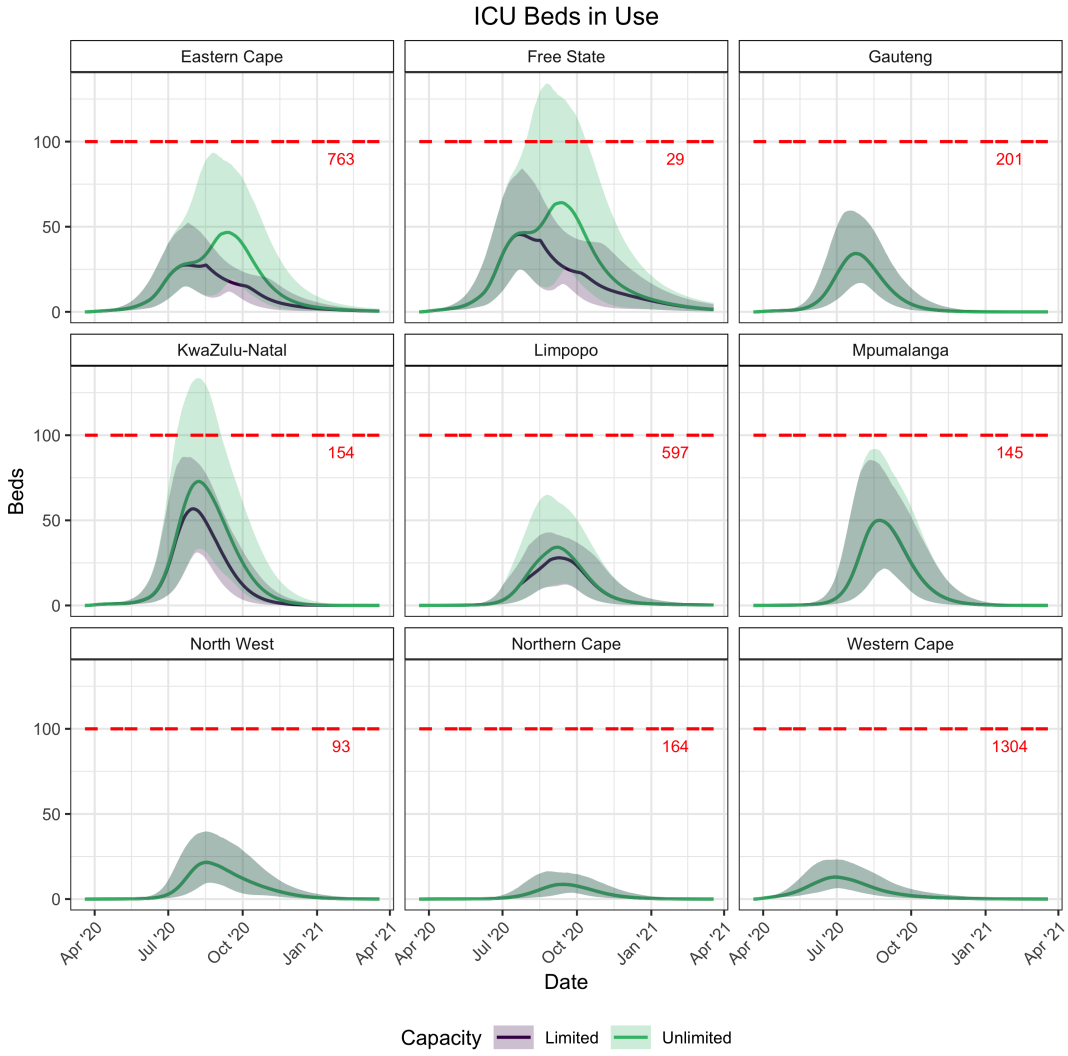

FIGURE 5.3: ICU beds in use over time per province. The red dashed lines represent capacity. The bands represent approximate 95% confidence intervals.

5.1.4 Deaths due to the non-availability of ICU beds constituted as much as 60% of total mortality

With the effect of capacity limitations on beds in use established, the next key question which needs to be addressed is the extent to which mortality due to the non-availability of ICU beds contributed to mortality in general. Deaths due to the non-availability of ICU beds are calculated as the total deaths occurring in waiting chambers. Figure 5.5 depicts mortality due to the non-availability of ICU beds as a proportion of total mortality. A reasonably large number of districts had non-zero proportions of mortality attributable to ICU overflow. The Eastern Cape and KwaZulu-Natal appeared to be most drastically affected by a lack of ICU beds, with proportions approaching sixty percent in the latter case. However, districts with high proportions of such mortality in these two provinces were often adjacent to districts with negligible mortality. This

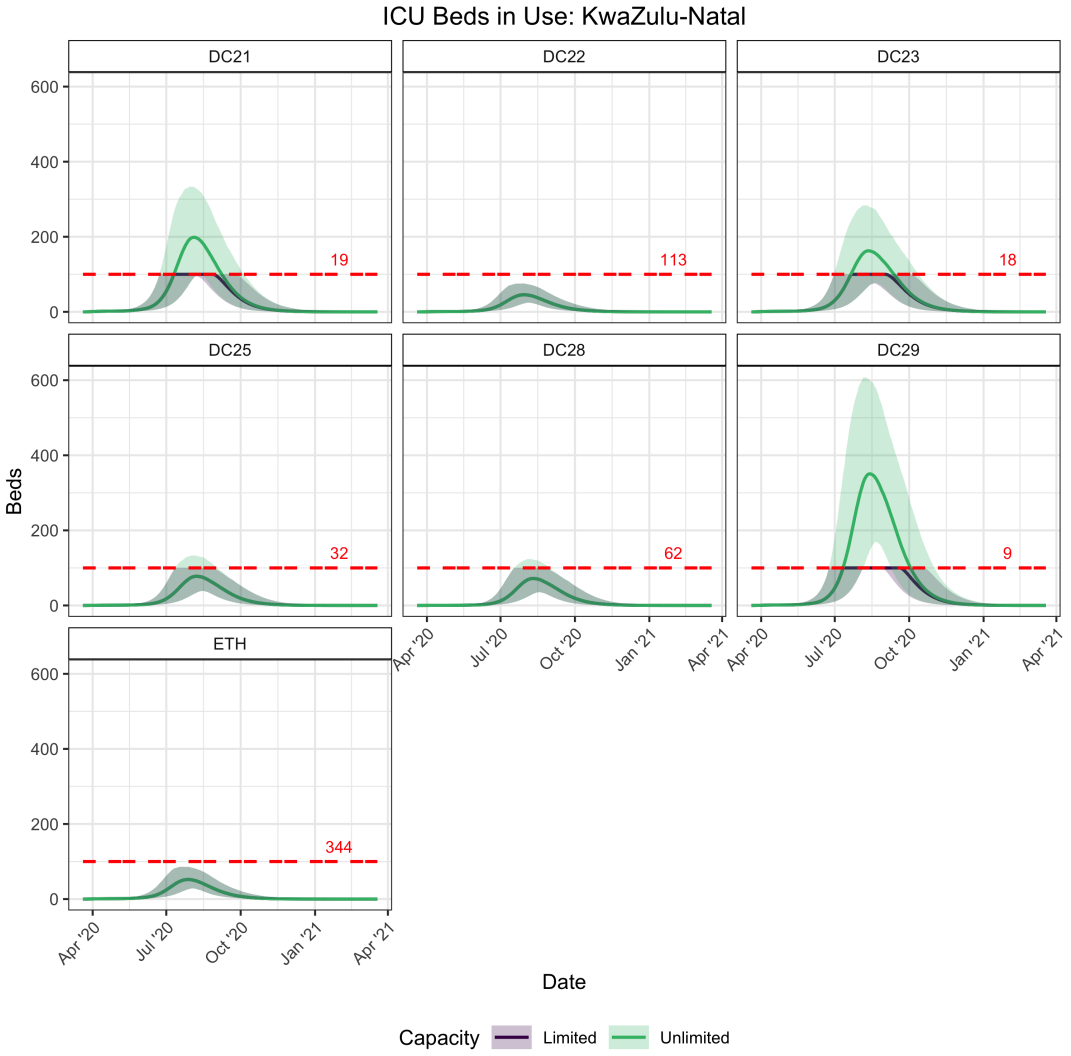

FIGURE 5.4: ICU beds in use over time per district within KwaZulu-Natal. The red dashed lines represent capacity. The bands represent approximate 95% confidence intervals.

disparity seems to be the largest in the worst case scenario; that is, the scenario representing the upper bound of an approximate 95% confidence interval. This further highlights the importance of intra-provincial resource sharing.

5.1.5 Deaths due to the non-availability of ICU beds decreased monotonically with increasing number of ICU beds

The extent to which mortality due to the non-availability of ICU beds contributed to overall mortality has been determined to be large, and the next step involves linking this to the goal programming approach. The goal programming framework works by shifting ICU beds around; hence, the relationship between the number of ICU beds in each district and the total number of deaths due to the non-availability of ICU beds must be explored. Figure 5.6 shows this relationship. Clearly, deaths monotonically decreased as the number of ICU beds increased. For all provinces, there appears to be a point at which deaths reached zero and further increases in the number of ICU

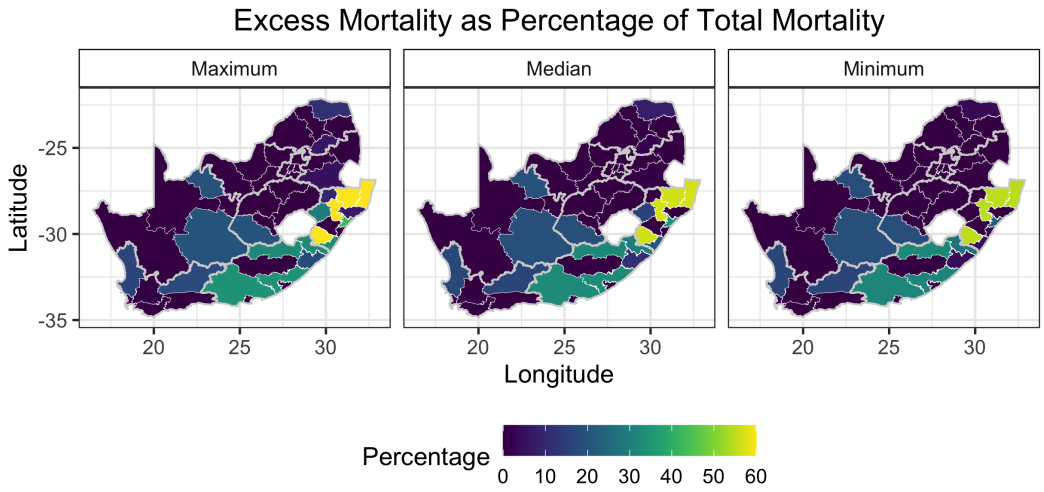

FIGURE 5.5: Mortality due to the non-availability of ICU beds shown as a percentage of total mortality, in the minimum, median and maximum cases. Minimum and maximum refer to the boundaries of an approximate 95% confidence interval.

beds has no effect. Gauteng and KwaZulu-Natal took the longest to reach this point. The Eastern and Western Cape provinces also took longer to reach zero than the rest of the provinces did. These disparities likely mirror what would be seen on a district level. Additionally, there appear to be the widest uncertainty bands around the curves of the provinces which reached zero at the highest numbers of ICU beds. The provinces for which the deaths reached zero at low numbers of ICU beds had very narrow uncertainty bands.

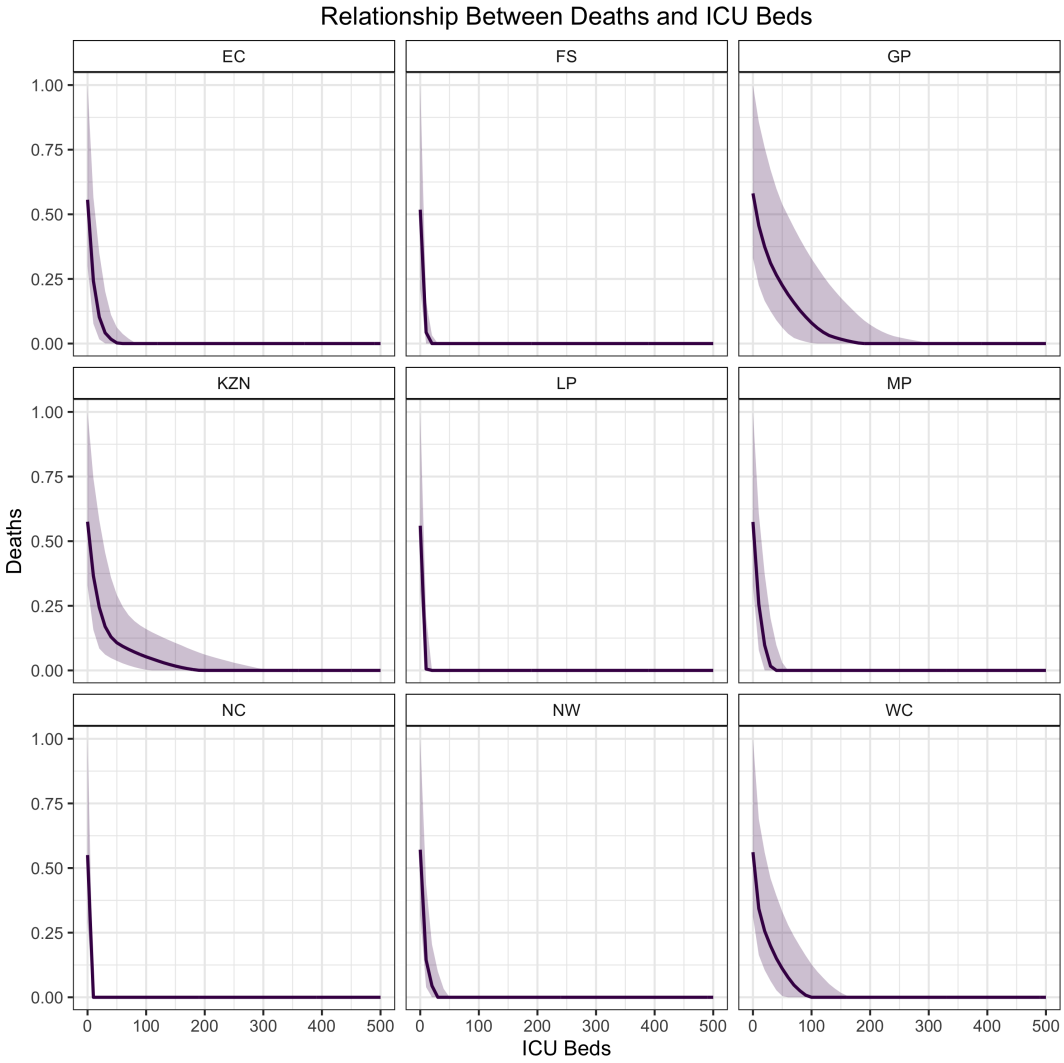

FIGURE 5.6: Relationship between number of ICU beds and number of deaths due to the non-availability of ICU beds. The bands indicate approximate 95% confidence intervals, which are simply the observed percentiles of the outputs. Deaths are expressed as the percentage of maximum deaths for each province, for better visibility of the complete curves.

5.2 Optimisation framework

The following key insights will be demonstrated in this section:

- Goal numbers of ICU beds varied across districts, but peaked at the peak of the epidemic.
- The largest shifts occurred around the peak of the epidemic, and some key districts were prominent in these shifts.
- Total cost was the highest for the beds-focused scenarios and lowest for the cost-focused scenarios.
- Total numbers of shifts were the highest for the beds-focused scenarios and lowest for the cost-focused scenarios.
- Total bed shortages were the highest for the cost-focused scenarios and lowest for the beds-focused scenarios.
- The goal programming framework reduced deaths due to the non-availability of ICU beds by 15% to 99%, depending on the scenario.
- Deaths were averted in all of the stochastic runs of the model under all scenarios.

5.2.1 Goal numbers of ICU beds varied across districts, but peaked at the peak of the epidemic

The first step in the goal programming and disease modelling framework is the generation of goals using the disease model. The results of this process are shown in Figure 5.7. In both the 95^{th} and 50^{th} percentile cases, the earliest and latest months' distributions peaked most sharply and were most right skewed. This indicates that in these months, especially in November and December, regardless of how conservative one chose to be, the goals for the majority of the districts were relatively low. This is because cases were projected to be at close to zero by December, as shown in Figure 5.1. The overall trend is that the distributions in both cases began right skewed with a very flat right tail, and steadily exhibited more blips towards the higher end of the goal spectrum as the midyear approached, after which the tails flattened again. Thus, as the months progressed from April to the middle of the year, there were more and more districts with goals closer to 300 ICU beds. This is intuitive, because the peak of the epidemic was experienced around July and August, which suggests that the need for ICU beds would have been the highest in those months. Aside from the behaviour of their right tails, the distributions for July and August also appeared to be bimodal and thus less peaked than the distributions for the earliest and latest months were. This indicates that the most variation in goals amongst the districts occurred in the middle of the year, which reflects variability in the predicted use of beds.

5.2.2 The largest shifts occurred around the peak of the epidemic, and some key districts were prominent in these shifts

With the goals generated by the disease model, the next step is to solve the goal programming problems and examine the corresponding solutions. The first aspect of this understanding involves looking at the pairs of districts between which resources were shifted, and the quantities of resources which were shifted. These relationships

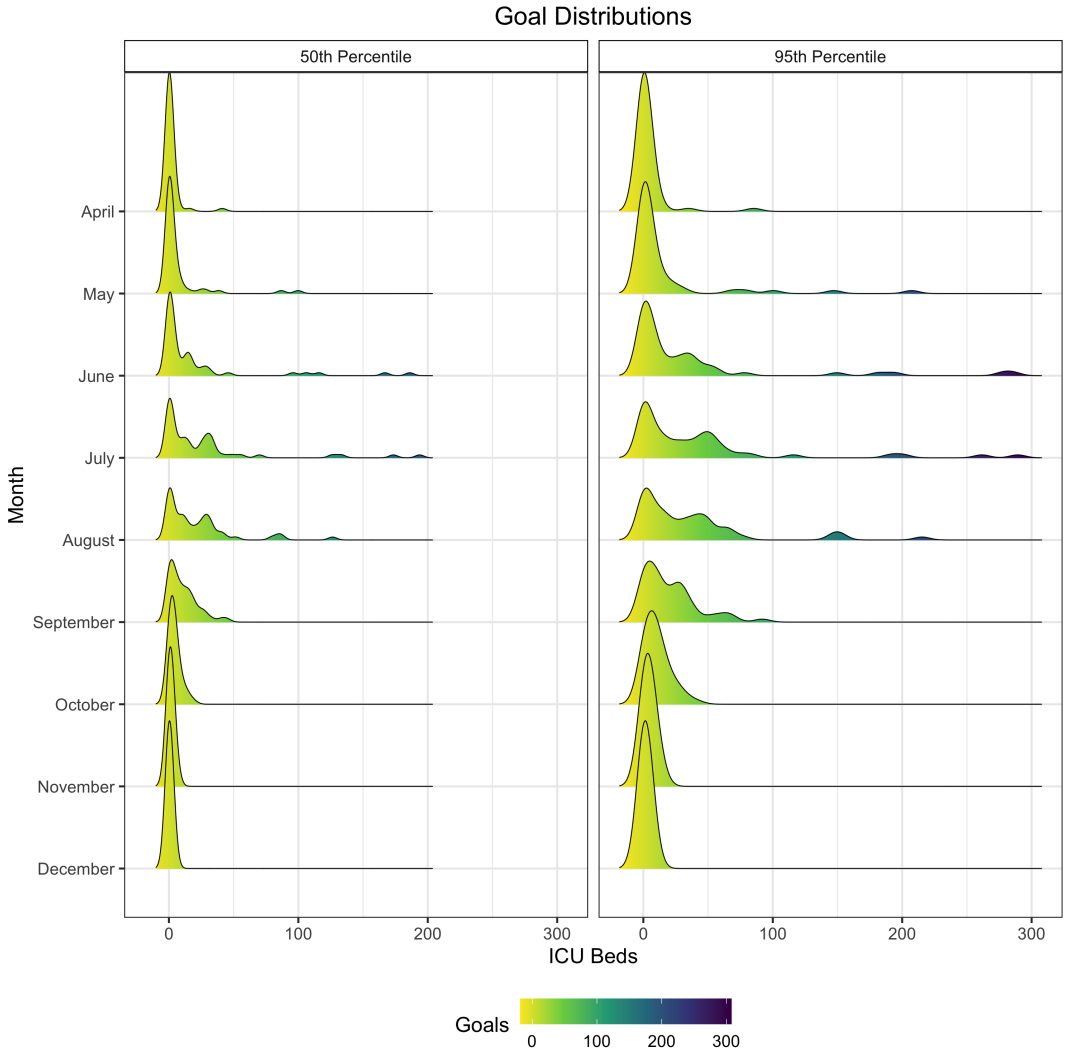

FIGURE 5.7: Monthly ICU bed goal distributions. These distributions are over the districts and are based on the 50^{th} (left) and 95^{th} (right) percentiles of maximum ICU beds needed in each district.

are depicted in Figure 5.8. There are several overall trends which are evident, as follows:

- With limited exceptions, the vast majority of shifts occurred from May to September. The latest that any shift occurred was in October, while in three cases, shifts occurred in April.
- The shifts that were largest in magnitude tended to occur in June, July, or August, which coincides with the peak of the epidemic and thus the time at which the goals would be most difficult to satisfy.
- In absolute terms, some of the largest shifts occurred between ETH and DC21, and ETH and DC29. Referring back to Figure 5.4 places this finding in context: DC29 was drastically exceeding capacity, while ETH did not exceed capacity, hence often shifting resources from ETH to DC29 makes sense. It similarly makes sense that ETH donated resources to DC21. Other pairings with large shifts between them were DC25 and DC26, DC28 and DC26, BUF and DC15,

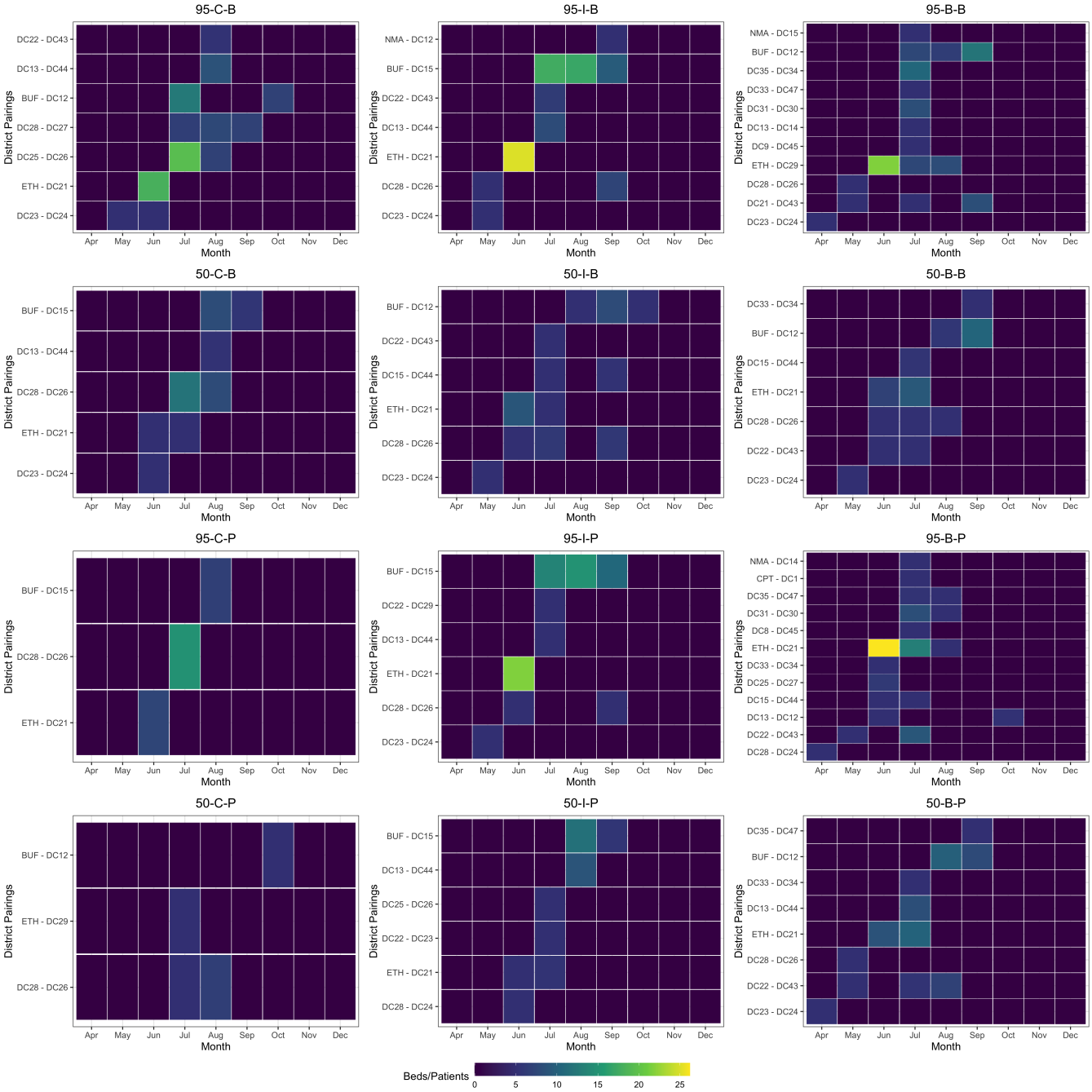

FIGURE 5.8: Numbers of resources shifted between each pair of districts for each goal programming scenario.

and BUF and DC12. DC25 and DC28 can also be seen in Figure 5.4, while DC26 was omitted due to its having zero capacity. It thus makes sense that resources would be shifted into DC26 from neighbouring districts such as DC25 and DC28.

- Finally, it is worth noting that scenarios 95-C-P and 50-C-P produced only three pairings, and these pairings included four of the same districts across the two scenarios. Scenarios 95-C-B and 50-C-B, on the other hand, produced seven and five pairings respectively. This may indicate that the weightings were a lot harsher in the case of patients than they were in the case of beds, in that they

forced cost minimisation to take more of a priority in the case of patients than they did in the case of beds.

5.2.3 Total cost was the highest for the beds-focused scenarios and lowest for the cost-focused scenarios

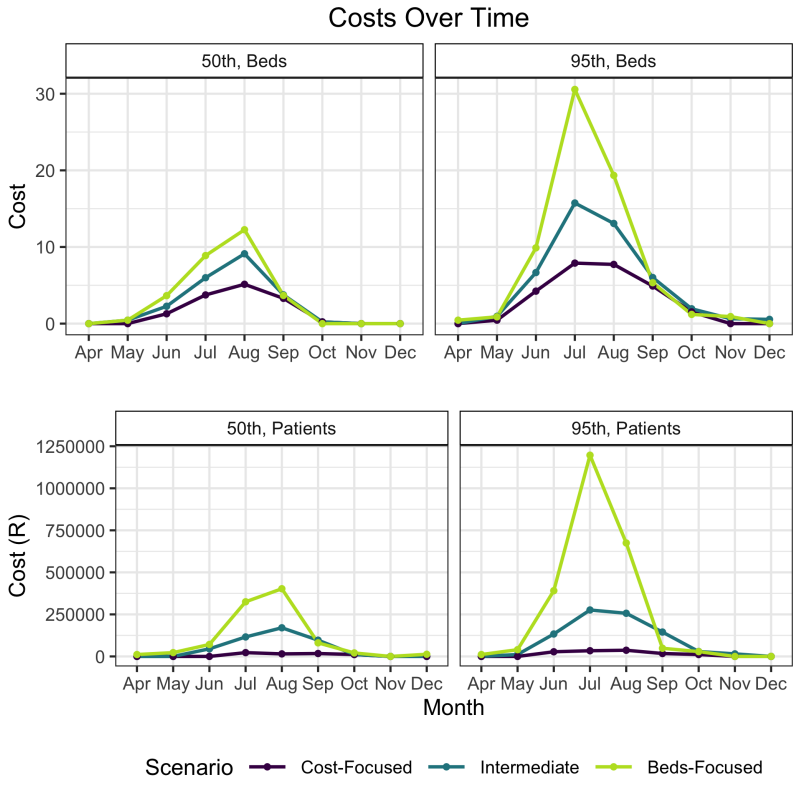

FIGURE 5.9: Total cost associated with each goal programming scenario, over time.

The next way in which a solution can be understood is by looking at the costs, numbers of shifts, and total shortages associated with each solution. Figure 5.9 indicates the total cost accumulated each month in each of the scenarios. There are several trends which are immediately evident. For most of the time, the beds-focused scenarios exhibited the highest costs, followed by the intermediate scenarios and then the cost-focused scenarios. This is as one would expect, as the beds-focused scenarios placed the highest emphasis on minimising shortages regardless of cost, while the cost-focused scenarios did the converse. In all cases, costs began at or near zero, and started to flatten out by October. When 50^{th} percentile goals were used, costs tended to peak in August, whereas they tended to peak in July when 95^{th} percentile goals were in use. This is because the more conservative 95^{th} percentile goals were higher earlier on in the epidemic. While the costs for shifting beds and patients were on very different scales, the patterns in cost curves were very similar; in both cases, the 95^{th} percentile curves tended to peak at higher numbers than their 50^{th} percentile counterparts did. Notably, the 95-B-P scenario peaked at a very high cost of just under R1,250,000, as opposed to its 50^{th} percentile counterpart, which peaked at less than R500,000. This represents more than double the cost when the goals were more conservative, which suggests that cost was highly sensitive to the level of the goals.

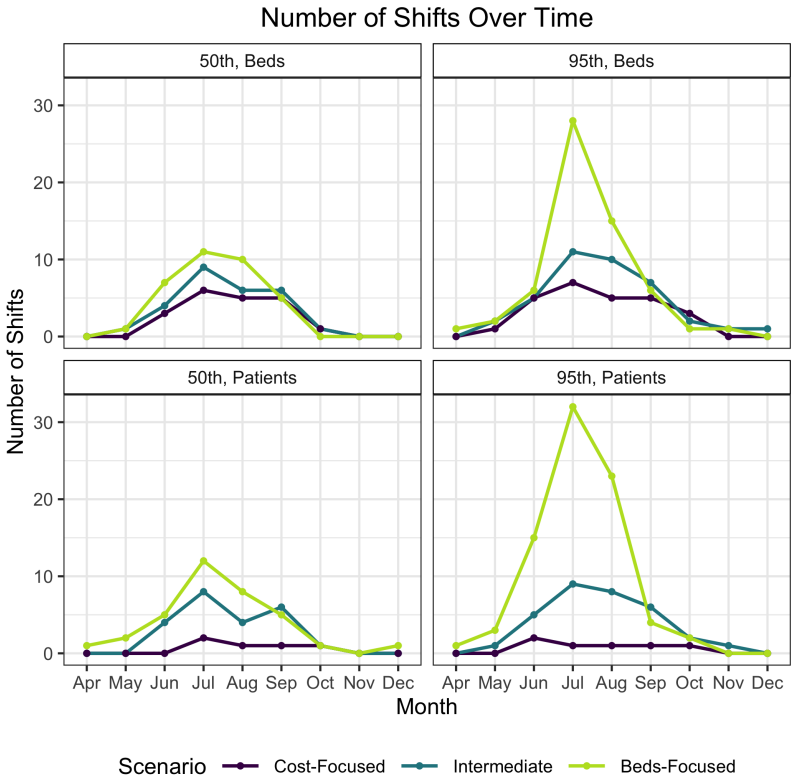

FIGURE 5.10: Number of shifts associated with each goal programming scenario, over time.

5.2.4 Total numbers of shifts were the highest for the beds-focused scenarios and lowest for the cost-focused scenarios

The total numbers of shifts occurring over time in each of the scenarios are shown in Figure 5.10. When 50^{th} percentile goals were in use, the numbers of shifts always peaked at just over ten, while when 95^{th} percentile goals were used, the peak numbers of shifts were closer to, if not in excess of, thirty. In both cases, peaks tended to occur in July, which once again coincides with the peak of the epidemic, and curves flattened out by September or October. The timing of these peaks coincides with the largest shifts seen in Figure 5.8. The flattening out of both the cost and the shortage curves mirrors the reduction in goals as the epidemic abated. For the most part, the cost scenarios exhibited the lowest numbers of shifts over time, regardless of the percentile of goals and of whether beds or patients were being shifted, followed by the intermediate scenarios and finally the beds-focused scenarios. This is both intuitive and in line with the findings in Figure 5.9; the higher the number of shifts in a particular solution, the higher the associated cost should be. While the costs were on different scales for beds and patients, the numbers of shifts are comparable. Figure 5.10 suggests that the solutions, at least in terms of number of shifts, were not particularly sensitive to the choice between beds and patients.

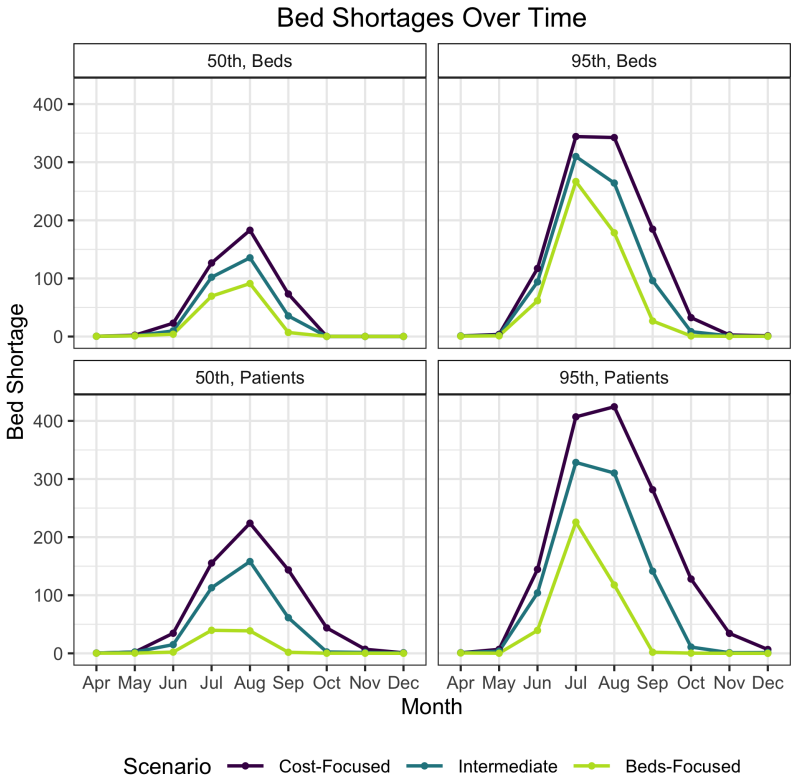

FIGURE 5.11: Total shortages associated with each goal programming scenario, over time.

5.2.5 Total bed shortages were the highest for the cost-focused scenarios and lowest for the beds-focused scenarios

The total shortages of beds in each of the scenarios over time are depicted in Figure 5.11. These shortages are defined as the differences between the respective goals and the available capacity figures in each month, where that difference is positive. Each month's available capacity is updated according to that month's shifts in each scenario. All of the curves followed a similar trend, in that there was an initial gradual climb, a sharp jump in May or June, a peak in July or August, and then a decline that ended in a flattening of the line. It makes sense that at the height of the epidemic, it would be most difficult to satisfy the goals, and as such, bed shortages would be highest. In general, the cost-focused scenarios exhibited the highest bed shortages, followed by the intermediate scenarios and then the beds-focused scenarios. This is once more intuitive, as the fewer shifts are made and the lower the associated cost, the higher the bed shortages are expected to be. Finally, the 50^{th} percentile scenarios produced shortages that peaked at approximately two hundred beds, whereas the 95^{th} percentile scenarios produced shortages which peaked at closer to, if not in excess of, four hundred. This is as one would expect: the higher the goals set, the higher the bed shortages would be expected to be.

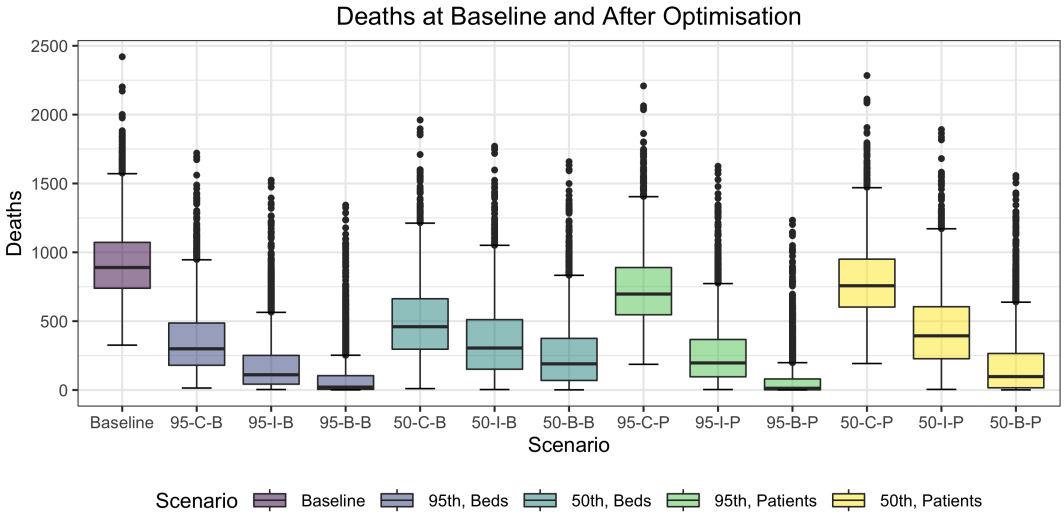

FIGURE 5.12: Deaths due to the non-availability of ICU beds, distributed over districts, before and after optimisation.

TABLE 5.1: Median deaths associated with each scenario, expressed in absolute terms and as a percentage of the scenario before optimisation takes place, along with associated total costs.

Scenario	Baseline	95-C-B	95-I-B	95-B-B	50-C-B	50-E-B	50-S-B
Median deaths	890	299	111	21	460	305	190
As % of baseline	100	34	12	2	52	34	21
Maximum	2420	1721	1524	1343	1961	1770	1658
Minimum	326	14	4	1	10	3	1
Cost		46	27	69	22	14	29
Scenario	Baseline	95-C-P	95-I-P	95-B-P	50-C-P	50-E-P	50-S-P
Median deaths	890	697	197	13	757	394	97
As % of baseline	100	78	22	1	85	44	11
Maximum	2420	2208	1625	1233	2284	1892	1558
Minimum	326	187	3	1	192	4	1
Cost (R)		1,180,349	697,264	1,762,642	566,698	363,716	750,862

5.2.6 The goal programming framework reduced deaths due to the non-availability of ICU beds by 15% to 99%, depending on the scenario

Finally, the results of the evaluation of the various goal programming solutions via the disease model are shown in Figure 5.12, with the accompanying median deaths and total costs shown in Table 5.1. Before any optimisation occurred, the median deaths was at 890, with the worst case scenario close to 2500. In all cases, the cost-focused scenario produced the highest median number of deaths, followed by the intermediate scenario, while the beds-focused scenario produced the lowest number of deaths. This is exactly as one would expect. When beds were shifted, the best of the cost-focused scenarios were observed, at 299 and 460 deaths when 95^{th} percentile and 50^{th} percentile goals were used, respectively. In the worst case scenario, when patients

were shifted, median deaths was close to or approximately 750, which represents more than a fifty percent increase when compared to shifting beds. The lowest numbers of deaths were observed in the 95^{th} percentile case, when beds or patients were shifted, with 21 and 13 median deaths respectively. These corresponded to the most expensive scenarios, at 69 units and R1,762,642 respectively. The 50^{th} percentile scenarios generally produced higher median deaths than their 95^{th} percentile counterparts did, suggesting a level of sensitivity of deaths to how conservative the goals were. This in turn means that it was prudent to be more conservative in setting goals. All of the distributions appear to be right-skewed; this skewness was most pronounced in the two best scenarios. It is worth noting that the worst data points in all of the scenarios approached, if not exceeded, 1500 deaths, which represents almost double the median deaths before any optimisation took place. It is therefore critically important to compare data points directly before and after optimisation.

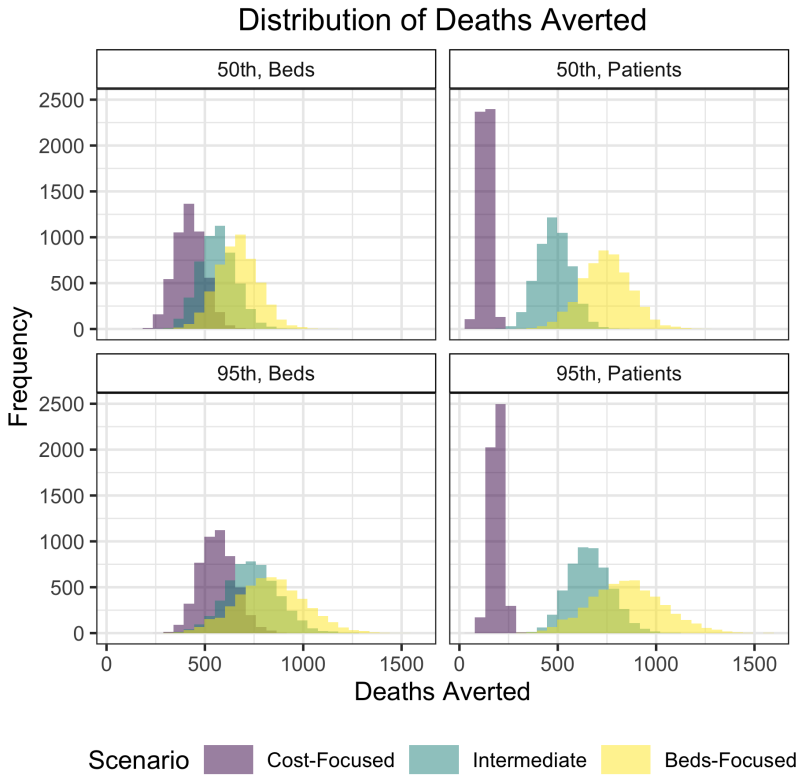

FIGURE 5.13: Distribution of total national deaths (due to the non-availability of ICU beds) averted due to the optimisation framework, over the 5000 simulation runs.

5.2.7 Deaths were averted in all of the stochastic runs of the model under all scenarios

This direct comparison, or rather, the deaths due to the non-availability of ICU beds that were averted due to the optimisation framework, are shown in the histograms in Figure 5.13. It is immediately evident from the positions of the histograms above zero that in no case did deaths due to the non-availability of ICU beds increase as a result of the shifts. Furthermore, all of the distributions were reasonably symmetric. When beds were shifted, there was a lot of overlap in the different weighting scenarios, with

the cost-focused scenario peaking at a higher value than in the case of the other two scenarios, followed by the intermediate scenario and finally the beds-focused scenario. In contrast, there was less overlap when patients were shifted, and particularly when 50^{th} percentile goals were used. The cost-focused scenario's distribution was shifted to the left, indicating fewer deaths averted, and was significantly more peaked than any of the other distributions were. This suggests that when high emphasis was placed on minimising cost, fewer lives were saved, which makes sense. The overlap between the intermediate and beds-focused distributions indicates that a reasonable amount of weighting could be placed on cost without too large of a sacrifice in terms of deaths averted, which bodes well for the use of this system in reality. Similarly, the large amount of overlap amongst all three scenarios in the case of shifting beds indicates that a poor choice of weightings could sometimes lead to a better outcome than a well-chosen weighting configuration would.

5.3 Conclusion

The results of the disease modelling highlighted the need for intra-provincial resource sharing, both in terms of the contribution of deaths due to the non-availability of ICU beds to overall mortality, and in terms of the exceeding of ICU bed capacity. Across all provinces, there was a clear relationship between the number of ICU beds and deaths due to the non-availability of beds, which provided sound justification for the goal programming approach of minimising bed shortages in order to minimise loss of life indirectly. Throughout all of the optimisation results, there was a clear pattern. As the epidemic peaked, so did the numbers of shifts and therefore the associated costs, and additionally the shortages. The district pairings plots, together with the ICU beds in use figures, painted an intuitive picture of intra-provincial resource sharing. Finally, the framework produced promising results. The lowest deaths overall were produced by the 95-B-P scenario at a cost of approximately R1,800,000, although the deaths averted distributions overlapped to a large extent when beds were shifted. When patients were shifted, the intermediate and beds-focused scenarios produced similar distributions.

6

Discussion

This chapter will review the results of this thesis in light of the context in which it finds itself, as well as in light of its various assumptions and limitations, the literature that came before it, and the future works which may build on it. It will further outline the value of this work.

The modelling results showed a clear need for intra-provincial resource sharing, and perhaps even inter-provincial sharing; however, the political backdrop against which these shifts would occur in reality must be acknowledged. Wealthier provinces might not be willing to pick up the slack of less well-resourced provinces, as even wealthy provinces have an array of needs competing for shares of a single budget which - in the developing world - is always less than ideal. This idea is exemplified in some of the findings in this thesis. The result that eThekwini (ETH) would have to shift large amounts of resources to neighbouring districts might be opposed, considering that eThekwini itself was very close to full capacity. Again, although neighbouring districts such as iLembe District Municipality (DC29) were in much more dire need of resources, shifting beds away from eThekwini might still cause some death, and it is difficult to maintain a utilitarian stance on the ground when this is the immediately visible cost. The case of eThekwini is perhaps an example of where inter-provincial shifting might stand a better chance of being realised, considering resources were spread so thinly in KwaZulu-Natal. This also highlights the need for provincial or national governmental control of health care resource allocation in the context of an epidemic.

At the very least, this framework provides some guidelines as to which districts should try to spare resources where they can, and which districts should be prioritised when there are spare resources available. The results further demonstrate the national advantages of co-operation amongst districts, which can be a powerful incentive for that co-operation to happen in reality. It would be very valuable to test the sensitivity of deaths averted to adherence to the schedules of shifts, and to establish whether using the schedules as loose guidelines rather than as a strict regimen would have any adverse effects. This presents a fascinating angle for future work. Another aspect of resource shifting that could be improved upon in future work is the allocation of resources below a district level: ideally, there should be another layer to the optimisation problem which looks at how to allocate resources amongst health facilities.

Although sensitivity analysis was somewhat built into the framework itself, this was not exhaustive and it is worth both noting this as a limitation of this thesis, and discussing the variables for which sensitivity analysis was not conducted. Firstly, increasing the minimum allowable shift makes the fifth constraint stricter, which means beds or patients are shifted in fewer cases. This would likely decrease cost but result in more loss of life, which presents an ethical issue. On the other hand, decreasing the minimum allowable shift results in the converse situation in which resources are shifted more frequently at a higher cost. The maximum allowable shift could theoretically be increased infinitely, which would mean that large quantities can be shifted, but this is bound by what is realistic in terms of both cost and logistics. This could also cause a situation in which large amounts are shifted to certain areas, allowing a higher proportion of deaths in others, which again is ethically questionable. That said, in the context of this application, increasing the maximum allowable shift would not likely have an effect because the shifts have not approached 1000. Decreasing the maximum shift enough would indeed have an effect and may result in certain districts' hoarding unused resources, which again is an ethical concern. Finally, increasing the cost parameters would make it more difficult to minimise cost, meaning that fewer shifts would be allowed and thus fewer lives will be saved.

This thesis is centred around the notion of capacity, a notion which must be revisited with the real context of an epidemic in mind. It is worth exploring both the rigidity of capacity constraints in reality, and the experience of South Africa particularly in the second wave of COVID-19. The estimates of available capacity in this thesis were meticulously calculated to reflect a realistic measure of what would be available on the ground; however, in reality, doctors often stretch capacity by finding innovative ways to take on more patients. In fact, the context discussed in Chapter 1 speaks directly to this, where in Houston, Texas, patients ended up being treated in emergency rooms, and London's Northwick Park's patients were tended to in operating theatres. As such, capacity constraints tend not to be hard constraints in reality, meaning that this thesis is perhaps premised on a slightly more pessimistic view of the epidemic than what actually happened. This means that the results of the framework could be even better when applied to the real world, where more capacity is available. However, this stretching of capacity is far too variable and unpredictable to be measured and taken into account. While around the world capacity was stretched by healthcare workers in the first wave, in many places the second wave presented a more dire situation. In South Africa in particular, the second wave was far larger than the first, pushing demand beyond what doctors could free up through innovation. Although this thesis was based on the first wave of COVID-19 in South Africa, the second wave context only increases its relevance and potential usefulness.

The question of which resources to shift was central to this thesis, and is worth discussing in some detail. The decision to look at patients and ICU beds was based on the availability of data. However, ICU beds are only valuable insofar as they are accompanied by mechanical ventilation and ICU nurses to administer it. So in reality, shifting ICU beds should be limited by the available supply of these resources at the destination. There are further oxygen-related considerations when it comes to shifting patients via ambulance. An oxygen tank only lasts one and a half hours, while high flow nasal oxygen in particular requires a high supply. A more detailed costing of shifting patients should incorporate oxygen tanks. Furthermore, there is a higher chance of mortality during transport; the prognosis worsens the longer the drive, suggesting

that transporting patients might be less feasible in larger provinces. There is precedent for transporting COVID-19 patients via ambulance; for example, patients were transported from Livingstone Hospital to a field hospital in the Eastern Cape. This was, however, a relatively short distance. This further supports the decision to limit resource shifting to be intra-provincial, and presents an opportunity for future work in which shifts are only allowed within a certain radius which is determined to be safe for transport. The unfortunate consequence of this reality is that rural areas would probably be at a disadvantage, due to their reduced proximity to resource-rich urban areas.

It is worth considering the intricacies of shifting health care workers instead of patients as a possible extension to this thesis. Health care workers do not require ambulances or oxygen, making them an appealing alternative. If on contract, health care workers are easy to shift from an administrative perspective; however, those employed by provincial Departments of Health would be difficult to shift inter-provincially. In general, embedded government staff would be more difficult to move. This is besides the fact that shifting health care workers requires uprooting people with established homes and families, and asking them to relocate for the duration of the epidemic is a large request. Asking them to shift around on a weekly basis is even more of an imposition, and would likely result in increased pressure and contribute to the exhaustion of the epidemic context. All of this holds in an ideal world, but in South Africa's context, where many nurses have gone on strike due to a lack of PPE - notably in the Eastern Cape - one would call into question the willingness of nurses to move from an overstaffed region to an under-staffed region. Although the idea of shifting nurses is at first glance quite attractive, these points highlight the difficulty thereof, and further justify the decision to focus on shifting patients and ICU beds in this thesis.

The results of shifting both ICU beds and patients rest somewhat on the cost assumptions of this thesis. The costs of shifting ICU beds were arbitrary proportions of distance, which meant that they could not be understood in context, nor could they be compared against those of shifting patients. This represents an area for improvement in future work. Moreover, while the cost of shifting patients was calculated in Rands, it did not take into account opportunity cost, which represents another interesting angle for future work. It is widely known that considering the impact of lockdowns, along with the redirection of resources towards COVID-19, saving each life from COVID-19 came at and still comes at a significant cost in terms of other diseases and other uses for ICU beds. Although the estimates of ICU beds available used in this thesis were based on the numbers that were expected to be available for COVID-19 use rather than the numbers of ICU beds available in absolute terms, one could still argue that basing a schedule of shifts purely on anticipated COVID-19 need could come at a high cost. Perhaps an extension to this thesis could also allow for fluctuations in other demands on ICU beds, such as trauma cases, which tend to peak in certain areas around Christmas time and Easter, due to road accidents and increased alcohol consumption.

The once-off nature of the framework - in that the disease model was run at the beginning of the year for the whole year, rather than monthly - warrants discussion. Ideally, one would run the model for one month ahead and determine shifts based on those runs, and then update the trajectory of the disease based on what actually happened, and run the model for another month ahead, and so on. This would ensure that the

goal programming would be working with the most recent and accurate information at each point in time. This could not, however, be implemented due to a lack of data. District-level case data were not necessarily available from month to month, while hospital admissions data were only accessible at the end of the first wave. Although this inability to rely on real-time data poses an unavoidable limitation to the framework, it also means that the performance of the framework was likely under-estimated, and may well be better were it to be provided with updating information. In any case, if one were to apply the framework as is in a real situation, its once-off forward looking plan could still act as a baseline plan from which to work, even in the presence of more recent information. In the absence of reliable real-time data, the computationally infeasible alternative is to follow multiple potential trajectories of the epidemic, and optimise over these trajectories. This could potentially be addressed via a dynamic programming approach, and presents a prospective topic for future work. It is worth noting that at the time of submission of this thesis, district level COVID-19 data are still not publicly available, which speaks to the challenges faced by many low and middle income countries (LMICs) in the context of a pandemic. This thesis thus represents a good practical solution for planning in the absence of detailed data.

Reporting of data is, in fact, an issue which necessitates dedicated discussion in and of itself. Of all provinces, only the Western Cape had one hundred percent of hospitals reporting cases in the first wave, but even in that case it did not translate to a situation in which one hundred percent of cases were reported. There are many reasons as to why under-reporting occurs in the midst of an epidemic. There were many people who were admitted and, especially at the height of the first wave, whose test results took so long to arrive that by the time they would have been diagnosed with COVID-19, they would have already died or been discharged, and thus escaped being recorded. Additionally, administrative systems - which are not perfect at the best of times - cannot be expected to function flawlessly in a crisis situation. For health care workers on the ground, filling in forms is far lower down on the priority list than actually tending to the surplus of patients is. Furthermore, many isolation cases in South Africa were in places like bed-and-breakfast accommodation rather than hospitals, meaning that they would not have been reported under hospital admissions. These factors mean that any analysis reliant on COVID-19 data should be interpreted with caution; this thesis is no exception.

It must not be forgotten that disease modelling is a simplification of the real world that rests on many levels of assumptions. Compartmental modelling in particular makes the assumption that the population is homogeneous except for disease and spatial state and that mixing is also homogeneous. This can lead to an overestimation of the total population infected (Bansal, Grenfell, and Meyers, 2007). In this model, heterogeneity in population behaviour has been taken into account through adjustments to the force of infection. The choice of which compartments to include is itself a simplification of the progression of the disease, and carries with it a host of assumptions. The various disease and treatment parameters are often unknown, more so with a novel virus, but are chosen and varied according to the best information available, often ruling out complex modelling methodology like agent-based models, which require incredibly granular data for model calibration. Despite these assumptions, compartmental modelling is arguably the most appropriate methodology available to synthesise data from all relevant aspects of the disease ecosystem into a computer-based simulation of health system dynamics. In the context of a pandemic

of a novel virus, compartmental modelling is also arguably a better choice than statistical data-driven modelling is, because the latter largely relies on sufficiently high quality data, whereas in a country like South Africa, data are often limited and of poor quality, as discussed. Additionally, the mechanistic nature of compartmental models allows for the projections of "what-if" scenarios, which are of high value to policy makers over and above the need for short-term forecasting.

The design and findings of this thesis must be placed in the context of the relevant literature. While many have modelled the spread of COVID-19, in the South African setting included, few have addressed the notion of limited capacity. Barasa, Ouma, and Okiro (2020)'s was the most comparable African study, which looked at estimating hospital and ICU bed surge capacity in Kenya's 47 counties. They found that there were significant geographical disparities in ICU bed resources, with only 22 counties' having at least one ICU unit. This is consistent with the findings of this thesis in South Africa. However, this thesis represents a more nuanced take on geographical disparities in resources in that it modelled infection in each of South Africa's fifty two districts, whereas Barasa, Ouma, and Okiro (2020) assumed a uniform proportion of symptomatic infection across all counties, based on a separate modelling study.

In the global context, Branas et al. (2020) aimed to estimate the excess mortality due to the non-availability of critical care beds in the USA. This closely matches one of the objectives of this thesis. Given the significant differences between the US and South African contexts, a direct comparison of findings is not possible; however, it can be observed that both this thesis and Branas et al. (2020) found geographical disparities in excess mortality. Branas et al. (2020) found that the highest excess mortality was concentrated in the New York City area, while there were seven other clusters of high mortality across the country. The greater Johannesburg region can be likened to the New York City area, and this thesis did not find a particularly high proportion of excess mortality in this region, likely because it is relatively well resourced compared to other districts. Consistent with the US study, however, this thesis found that excess mortality due to the non-availability of ICU beds was the biggest contributor to overall mortality in one particular area: that is, KwaZulu-Natal. Overall, this thesis produced well aligned modelling results when compared to the literature, and represents a valuable contribution to South African research in particular.

In the process of any health economic analysis, the question of costing human life is inevitable. In this thesis, the Rand value a decision maker is willing to give up in exchange for a bed in the case of shifting patients could, to the decision maker, practically translate to how much a life is worth. While some might assert that the moral stance is that a human life is priceless - or perhaps that human cost does not have a financial equivalent - the reality is that governments and other decision makers have long been valuing human life whenever scarce resources have been allocated, especially in the medical arena. The COVID-19 pandemic has simply brought these valuations into the realm of public discussion. While there is no easy answer, it is important both to bear in mind the sensitivity of the topic, and to take it in context. This thesis asks the question of a very specific budget allocation that in reality would operate within fixed parameters. It does not ask the decision maker to value a human life in absolute terms, but rather to place as much value as possible on human life while remaining in the realms of what is realistic for a developing nation's budget. Being

cognisant of this underlying fact is necessary to grasp the gravity of the decision at hand.

Ultimately, the contributions of this thesis both to science and to the practical world of governmental decision making in the context of a global pandemic can be summarised in two areas. Firstly, a large portion of the work's value has been seen to fruition through the use of results by key South African stakeholders such as the National Department of Health, National Treasury, and various levels of government. Secondly, the work is a practical demonstration of integrating epidemic modelling and optimisation in the COVID-19 context. While disease modelling has been ubiquitous in the world's handling of COVID-19, this thesis shows the possible impact that resource constraints and optimisation of resources can have on an epidemic, using South Africa as a case study. It is easy to generalise the framework to the shifting of any resources in a health care context, such as mobile health clinics for the elimination of malaria in South Africa.

6.1 Conclusion

Despite its limitations and assumptions, this thesis has already added value to the South African health system in the time of COVID-19. It has further shown the value of intra-provincial resource sharing, and the impact that nationwide co-operation can have in the context of a pandemic. The framework that this thesis proposes provides guidelines for health systems like South Africa's, that are applicable both to general health care problems in a resource-constrained setting and to future pandemic preparedness.

7

Conclusion

This thesis had two primary aims centred around mitigating the effects of the first wave of the COVID-19 epidemic in SA. The first was fulfilled during the development of the NCEM, via the inclusion of capacity constraints and waiting chambers into the Model. The second was to develop a goal programming framework that would minimise loss of life due to the non-availability of ICU beds. This chapter will cover the fulfilment of these aims, as well as that of the more detailed objectives set out in Chapter 1. It will further highlight the key findings of, and conclusions that can be drawn from, this thesis.

A descriptive analysis of the four key data-sets used in this thesis established, amongst others, two important points of context:

- The high level of disparity in ICU bed supply across SA's districts made a preliminary case for intra-provincial sharing amongst districts.
- The peak of SA's first wave of COVID-19 - the focus of this thesis - occurred around July and August 2020.

This thesis modelled the spread of COVID-19 across SA's 52 districts using the NCEM, a generalised SEIR model which captures the clinical stages of the disease, as well as treatment pathways. The contribution to the NCEM that is attributable to this thesis is the addition of waiting chambers which capture people who cannot be treated once general hospital bed or ICU bed capacity has been reached. This contribution involved the development of a capacity constraint algorithm. The main insights from the modelling results were as follows:

- ICU bed capacity was projected to be breached in two provinces. KwaZulu-Natal was explored as a district-level example, and showed breaches in five out of eleven districts, highlighting the need for intra-provincial resource sharing.
- An examination of mortality due to the non-availability of ICU beds revealed that it constituted as much as 60% of total mortality in certain KwaZulu-Natal districts. This made it clear that deaths due to the non-availability of ICU beds was an outcome worth mitigating.
- These deaths decreased with increasing numbers of ICU beds across all provinces, which meant that aiming to minimise ICU bed shortages in each district should translate to minimising deaths due to the non-availability of ICU beds.

The disease model was then used to generate the goals - taken as percentiles of the maximum ICU bed usage over the period of a year - for a goal programming problem which aimed to minimise ICU bed shortages in each district, while also minimising cost. Twelve scenarios were designed and tested. The following key takeaways were established:

- Goals, bed shortages, costs of shifting, and numbers of shifts peaked around the peak of the first wave in SA.
- The goal programming framework reduced deaths due to non-availability of ICU beds by at least 15%, and up to 99%, depending on the scenario. The best scenario was 95-B-P, and the best scenario when shifting beds was 95-B-B. The costs of these scenarios were approximately R1,800,000 and 69 units, respectively.

Despite its limitations and assumptions, this thesis made a contribution to South African governmental decision making via its additions to the NCEM. The framework it proposes highlights the potential positive impact of province-wide or even country-wide resource sharing as the world faces further waves of COVID-19, and indeed future pandemics.

A

Proxy for Cost of Transporting Beds

FIGURE A.1: Scaled distances between districts, which represents a proxy for the cost of transporting beds.

B

ICU Beds in Use Per District

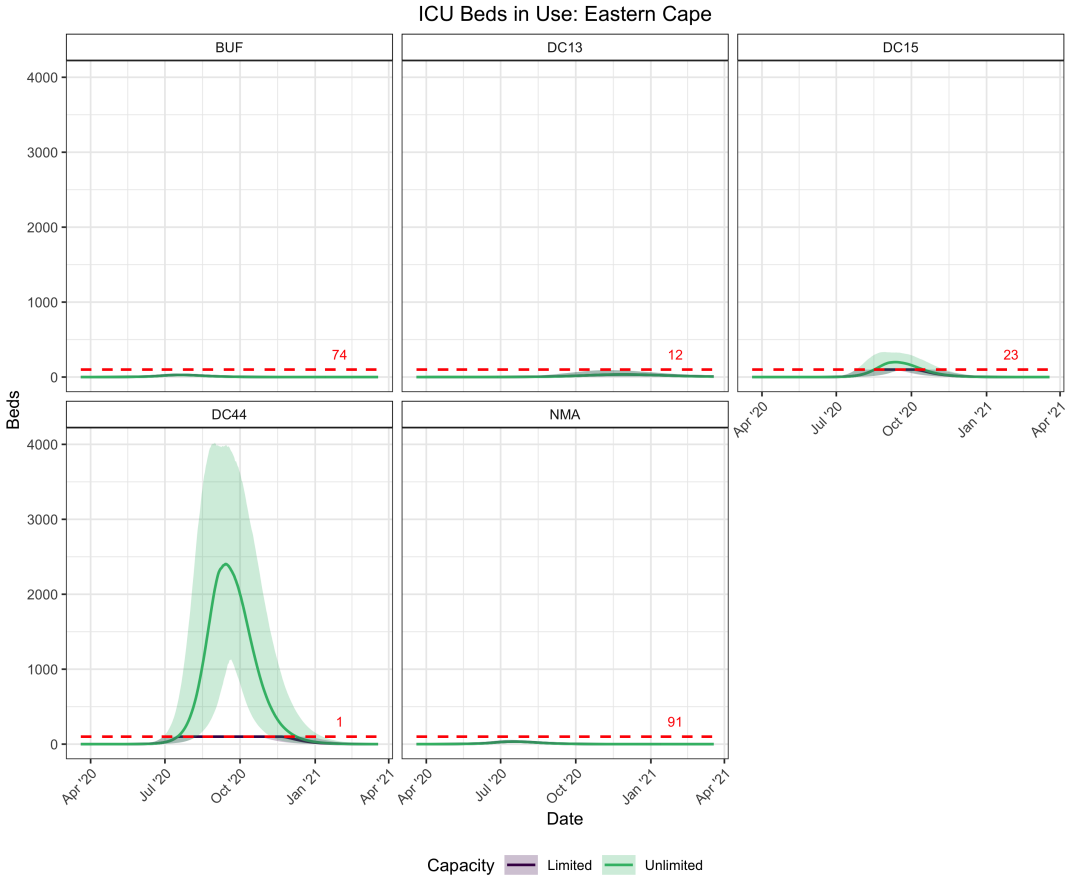

FIGURE B.1: ICU beds in use over time per district within the Eastern Cape. The red dashed lines represent capacity. The bands represent approximate 95% confidence intervals.

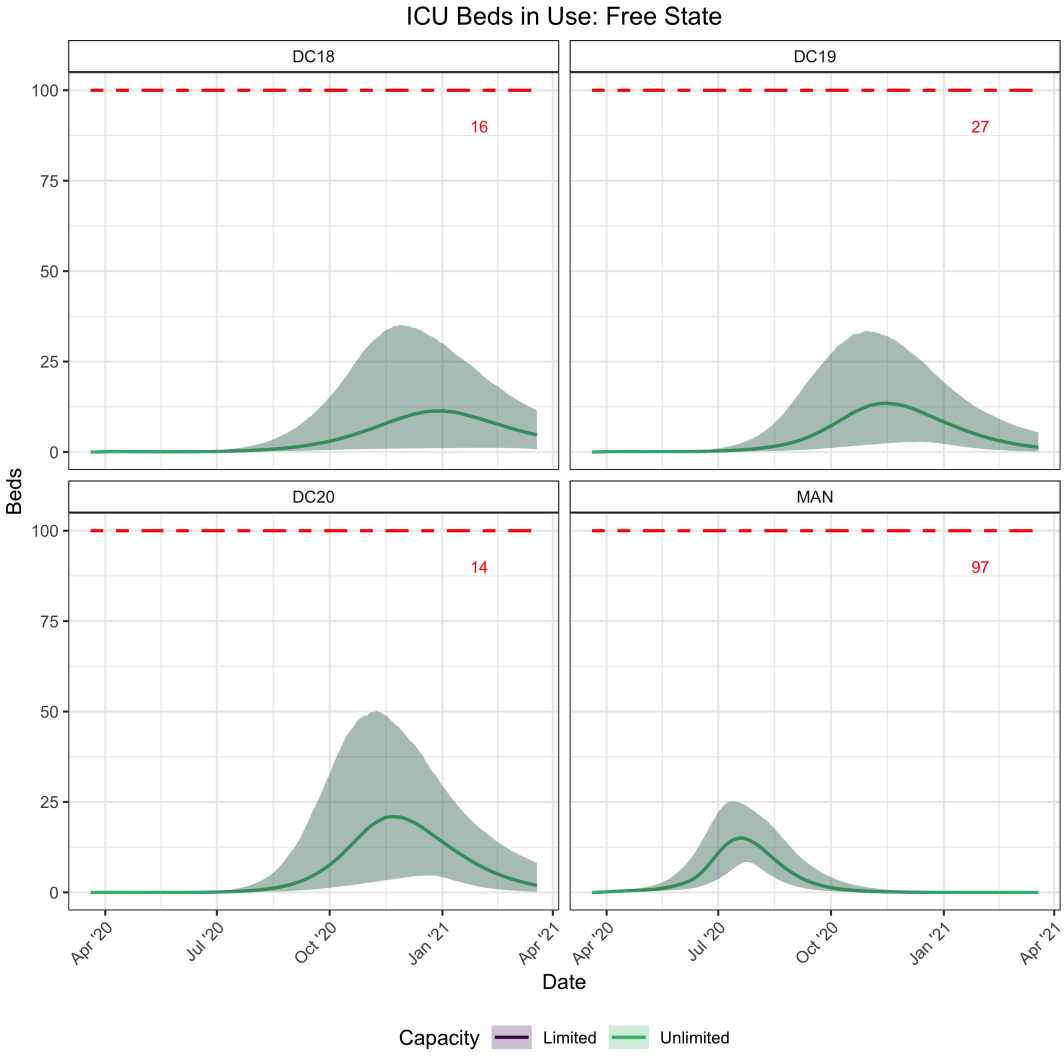

FIGURE B.2: ICU beds in use over time per district within the Free State. The red dashed lines represent capacity. The bands represent approximate 95% confidence intervals.

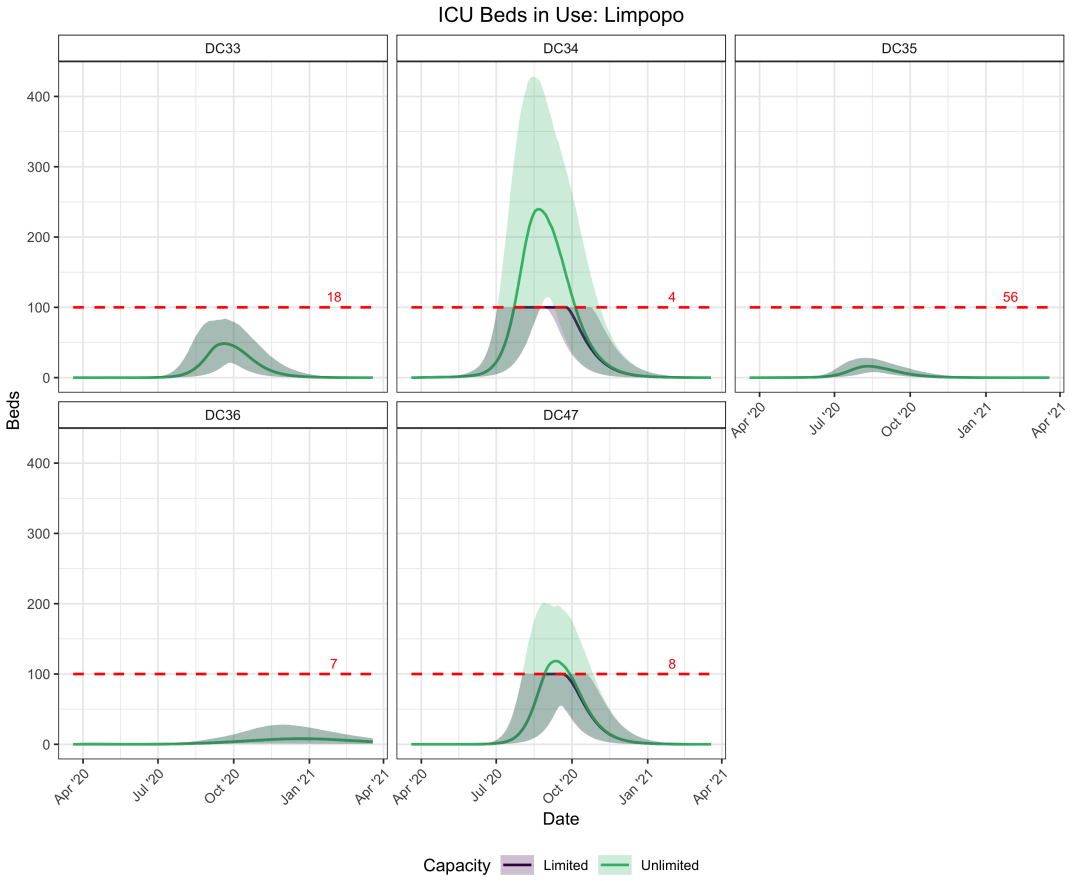

FIGURE B.3: ICU beds in use over time per district within Limpopo. The red dashed lines represent capacity. The bands represent approximate 95% confidence intervals.

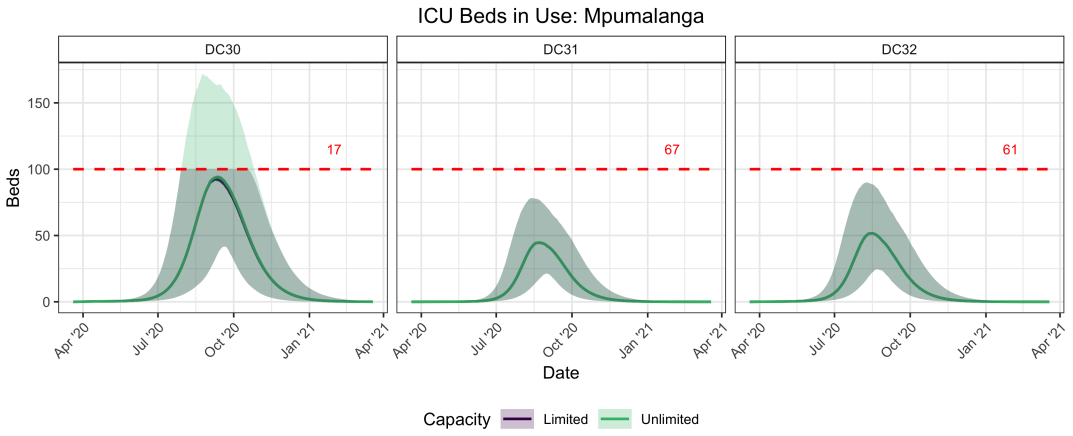

FIGURE B.4: ICU beds in use over time per district within Mpumalanga. The red dashed lines represent capacity. The bands represent approximate 95% confidence intervals.

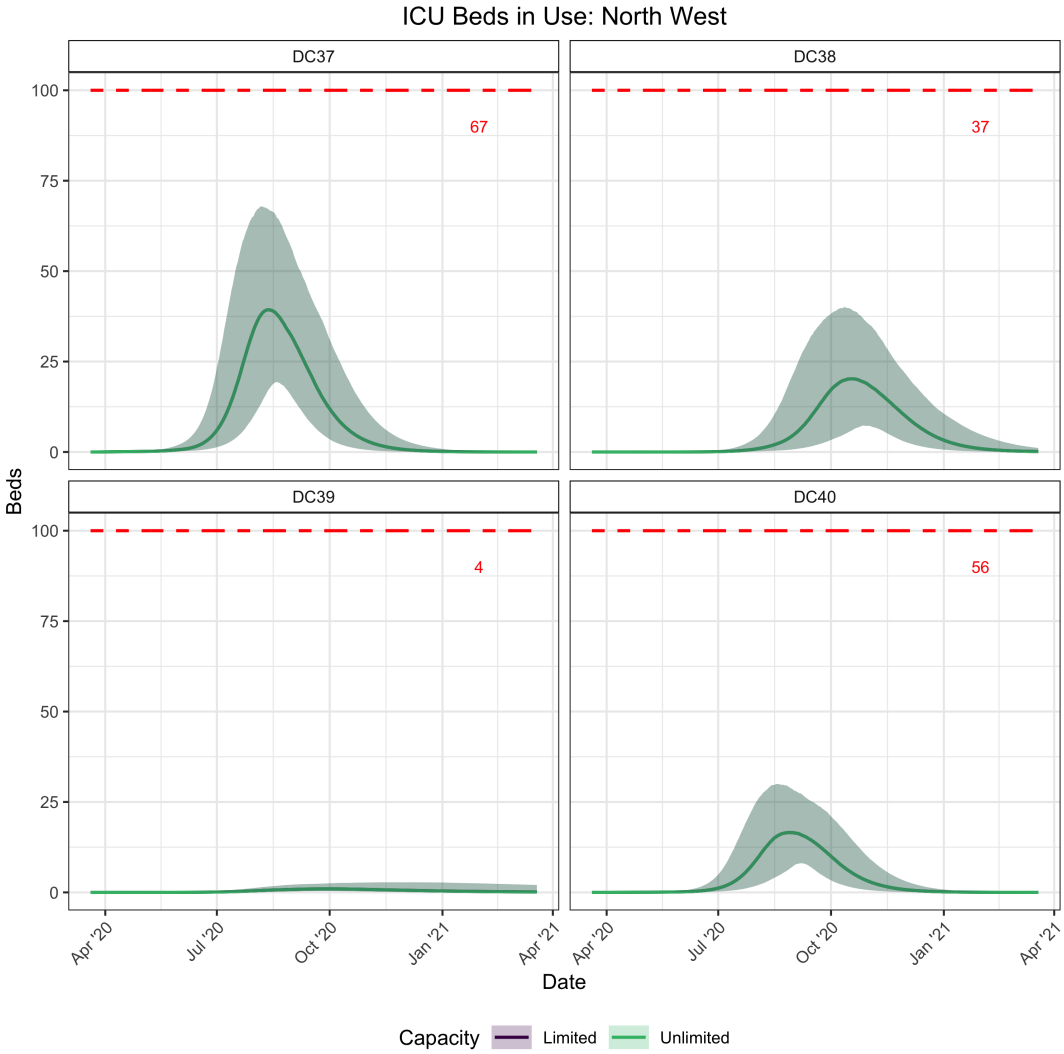

FIGURE B.5: ICU beds in use over time per district within the North West. The red dashed lines represent capacity. The bands represent approximate 95% confidence intervals.

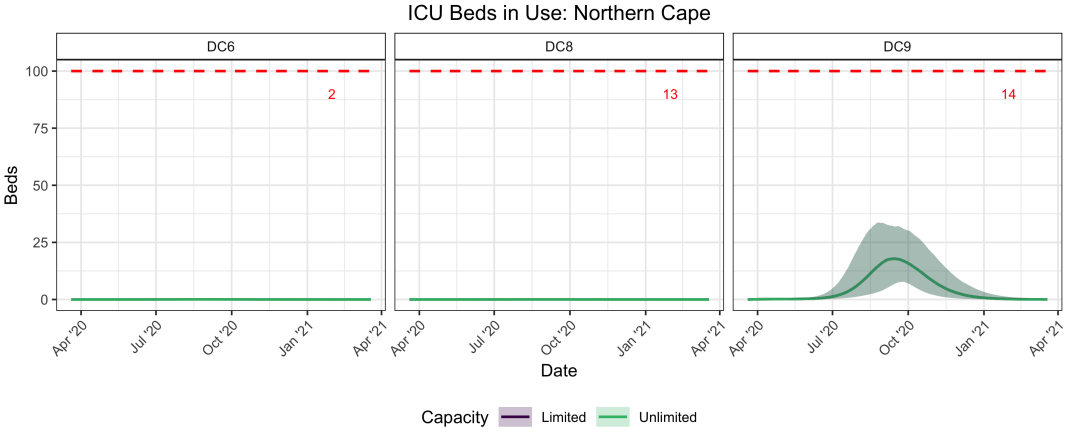

FIGURE B.6: ICU beds in use over time per district within the Northern Cape. The red dashed lines represent capacity. The bands represent approximate 95% confidence intervals.

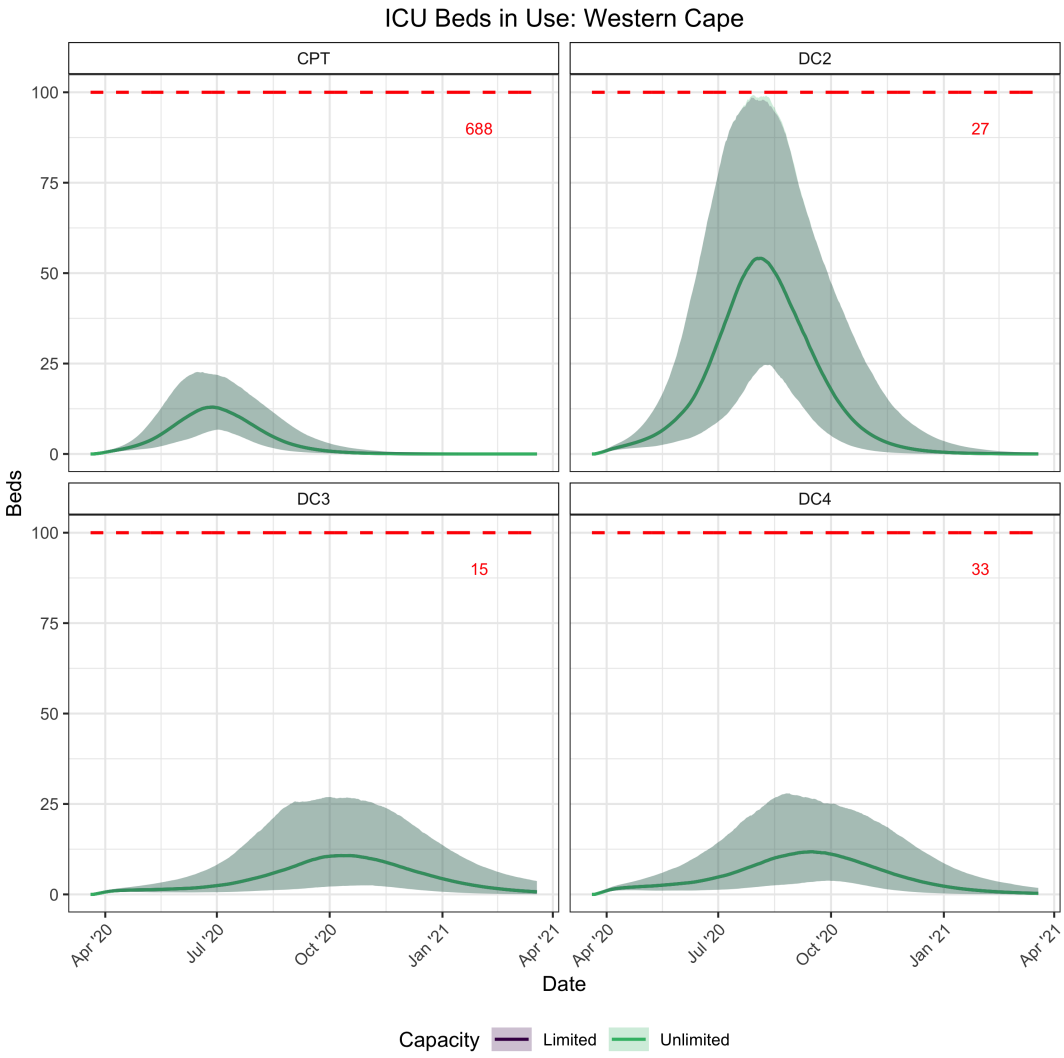

FIGURE B.7: ICU beds in use over time per district within the Western Cape. The red dashed lines represent capacity. The bands represent approximate 95% confidence intervals.

C

Selected Code Snippets

C.1 Capacity Constraint Algorithm

Note that the full code for the NCEM is available at https://sacovid19mc.github.io.

```
# Excess mortality
avail1 = rep(1, N)
avail2 = rep(1, N)

# Hospital beds
beds_occupied <- Reduce("+", lapply(c(7, 8, 11, 19, 20), function(ind){x[varind[ind,]]})) %>% unname
# Sev to H1 and H2
proposed_enter = unname(taus*x[varind[6,]]/365)
proposed_exit = unname(
            tauicu*x[varind[8, ]]+ # H2 to ICU
            r3*x[varind[7, ]]+          # H1 to R,D
            r5*x[varind[11, ]]+         # H3 to R
            r9*x[varind[19, ]]+       # Wv to R,D
            r10*x[varind[20, ]]       # WnotV to R,D
          )/365
violation = which(beds_occupied - proposed_exit + proposed_enter > gen_capacity)
avail1[violation] = (gen_capacity[violation] - beds_occupied[violation]
     + proposed_exit[violation]) / proposed_enter[violation]

# ICU
ICU_occupied <- Reduce("+", lapply(c(9, 10, 16, 17), function(ind){x[varind[ind,]]})) %>% unname
# H2 to all ICUs
proposed_enter = unname(tauicu*x[varind[8,]]/365)
proposed_exit = unname(
            muv*x[varind[9, ]] +        # ICUVd to D
            r9*x[varind[10, ]] +          # ICUVr to H3
            munotv*x[varind[16, ]] + # ICUnotVd to D
            r10*x[varind[17, ]]           # ICUnotVr to H3
          )/365
violation = which(ICU_occupied - proposed_exit + proposed_enter > icu_capacity)
avail2[violation] = (icu_capacity[violation] - ICU_occupied[violation]
    + proposed_exit[violation]) / proposed_enter[violation]

# Correct for zero-capacity situations
avail1[which(gen_capacity == 0)] = 0
avail2[which(icu_capacity == 0)] = 0
```

C.2 Goal Programming Solution

```
get_sol = function(scenario,    # 'anywhere' or 'neighbours' or '
    provincial'
                   add_constraint,    # TRUE or FALSE (deprecated)
                   shortage_weight,  # number from 0 to 1
                   cost_weight,       # number from 0 to 1
                   goals,               # vector of goals of length N
    =52
                   minshift=5,
                   maxshift=1000,
                   available=capacity$ICU_curr,
                   ems=FALSE) {                  # include EMS
    constraint

   library(Rglpk)

   # CONSTANTS ####
   epsilon = 0.001
   N = 52
   weights = c(rep(shortage_weight, N), cost_weight)
   if(ems == FALSE) cost = as.vector(t(distance / 1000000))
   else cost = as.vector(t(cost_ems))

   if(scenario == 'neighbours') {
     neighbours_vec = readRDS('Model Inputs/neighbours.RDS')
     neighbours_vec = as.vector(t(neighbours_vec))
   }
   if(scenario == 'provincial') {
      provincial_vec = readRDS('Model Inputs/provincial.RDS')
      provincial_vec = as.vector(t(provincial_vec))
   }

   # OBJECTIVE ####
   obj = c(1,
           rep(epsilon, N + 1) * weights,
           rep(0, 2*N^2))

   # GOALS AND CONSTRAINTS: MATRIX ####

   # Tchebycheff constraints
   ones = rep(1, N)
   ws = -diag(weights[1:N])
   zeros = matrix(0, nrow=N, ncol=2*(N^2) + 1)
   mat = cbind(ones, ws, zeros)
   mat = rbind(mat, c(1, rep(0, N), -weights[N+1], rep(0, 2*(N^2))))

   # Reaching goal capacity but staying within what's available
   iden = cbind(rep(0, N), diag(rep(1, N)), rep(0, N))
   i_to_j = matrix(0, nrow=N, ncol=N^2)
   for(i in 1:N) {
     for(j in 1:N) {
       i_to_j[i, i + (j - 1) * N] = 1
     }
   }
   for(i in 1:N) {
     i_to_j[i, (N*i - (N - 1)):(N *i)] = -1
   }
   mat = rbind(mat,
               cbind(iden,
                     i_to_j,
                     matrix(0, nrow=N, ncol=(N^2))))
```

```
58
59    # Cost goal
60    cost_vec = c(rep(0, N + 1), -1, cost, rep(0, N^2))
61    mat = rbind(mat, cost_vec)
62
63    # Minshift
64    zeros = matrix(0, nrow=N^2, ncol=N+2)
65    iden = diag(rep(-1, N^2))
66    ms = diag(rep(minshift, N^2))
67    mat = rbind(mat,
68                cbind(zeros,
69                      iden,
70                      ms))
71
72    # Maxshift
73    zeros = matrix(0, nrow=N^2, ncol=N+2)
74    iden = diag(rep(1, N^2))
75    ms = diag(rep(-maxshift, N^2))
76    mat = rbind(mat,
77                cbind(zeros,
78                      iden,
79                      ms))
80
81    # Shift away cannot exceed available
82    avail = matrix(0, nrow=N, ncol=N^2)
83    for(i in 1:N) {
84      avail[i, ((i - 1) * N + 1):(i * N)] = 1
85    }
86    avail = cbind(matrix(0, nrow=N, ncol=N+2),
87                  avail,
88                  matrix(0, nrow=N, ncol=N^2))
89    mat = rbind(mat, avail)
90
91    # Delta's zero for i=j
92    iden = matrix(0, nrow=N, ncol=N^2)
93    for(i in 1:N) {
94      iden[i, (N + 1)*i - N] = 1
95    }
96    iden = cbind(matrix(0, nrow=N, ncol=2+N+N^2),
97                 iden)
98    mat = rbind(mat, iden)
99
100   if(add_constraint == TRUE) {
101     # deltas zero when gi > ai
102     zeros = matrix(0, nrow=N^2, ncol=2+N+N^2)
103     shortage = goals[1:N] > available[1:N]
104     delt = matrix(0, nrow=N^2, ncol=N^2)
105     for(i in 1:N) {
106       if(shortage[i] == TRUE) {
107         diag(delt[((i - 1) * N + 1):(i * N), ((i - 1) * N + 1):(i * N)
    ]) = 1
108       }
109     }
110     mat = rbind(mat,
111                 cbind(zeros,
112                       delt))
113   }
114
115   if(scenario == 'neighbours') {
116     # Only shift to neighbours
117     iden = diag(rep(1, N^2))
118     zeros = matrix(0, nrow=N^2, ncol=N + 2 + N^2)
119     mat = rbind(mat,
```

```
                cbind(zeros,
                      iden))
  }

  if(scenario == 'provincial') {
    # Only shift provincially
    iden = diag(rep(1, N^2))
    zeros = matrix(0, nrow=N^2, ncol=N + 2 + N^2)
    mat = rbind(mat,
                cbind(zeros,
                      iden))
  }

  if(ems == TRUE) {
     # constrain shifts to number of ambulances
     avail = matrix(0, nrow=9, ncol=N^2)
     for(i in 1:N) {
        if(i %in% P$EC) avail[1, ((i - 1) * N + 1):(i * N)] = 1
        if(i %in% P$FS) avail[2, ((i - 1) * N + 1):(i * N)] = 1
        if(i %in% P$GP) avail[3, ((i - 1) * N + 1):(i * N)] = 1
        if(i %in% P$KZN) avail[4, ((i - 1) * N + 1):(i * N)] = 1
        if(i %in% P$LP) avail[5, ((i - 1) * N + 1):(i * N)] = 1
        if(i %in% P$MP) avail[6, ((i - 1) * N + 1):(i * N)] = 1
        if(i %in% P$NC) avail[7, ((i - 1) * N + 1):(i * N)] = 1
        if(i %in% P$NW) avail[8, ((i - 1) * N + 1):(i * N)] = 1
        if(i %in% P$WC) avail[9, ((i - 1) * N + 1):(i * N)] = 1
     }
     avail = cbind(matrix(0, nrow=9, ncol=N+2),
                   avail,
                   matrix(0, nrow=9, ncol=N^2))
     mat = rbind(mat, avail)
  }


  # GOALS AND CONSTRAINTS: RHS ####
  rhs = c(rep(0, N+1),
          goals - available,
          rep(0, 1+2*N^2),
          available,
          rep(0, N))

  if(add_constraint == TRUE) rhs = c(rhs, rep(0, N^2))
  if(scenario == 'neighbours') rhs = c(rhs, neighbours_vec)
  if(scenario == 'provincial') rhs = c(rhs, provincial_vec)
  if(ems == TRUE) rhs = c(rhs, c(349, 166, 513, 290, 144, 67, 77, 79,
   230))

  # GOALS AND CONSTRAINTS: DIRECTION ####
  dir = c(rep(">=", 2*N+1),
          "==",
          rep("<=", 2*N^2 + N),
          rep("==", N))

  if(add_constraint == TRUE) dir = c(dir, rep('==', N^2))
  if(scenario == 'neighbours') dir = c(dir, rep('<=', N^2))
  if(scenario == 'provincial') dir = c(dir, rep('<=', N^2))
  if(ems == TRUE) dir = c(dir, rep('<=', 9))

  # VARIABLE TYPES ####
  types = c(rep('C', N + 2),
            # rep('I', N^2),
            rep('C', N^2),
            rep('B', N^2))
```

```
182
183     # SOLUTION ####
184     sol=Rglpk_solve_LP(obj, mat, dir, rhs, types = types,
185                        verbose=T,
186                        control = list(tm_limit = 300000))
187
188     # PROCESS SOLUTION ####
189     shifts = sol$solution[(2 + N + 1):length(sol$solution)][1:(N^2)]
190     shifts = matrix(shifts, nrow=N, ncol=N, byrow=T)
191     dv = sol$solution
192     delta = dv[(N + 2 + N^2 + 1):(length(dv))]
193     delta = matrix(delta, nrow=N, byrow=T)
194
195     # RETURN ####
196     return(list(shifts=shifts, delta=delta))
197 }
```

Bibliography

Adam, D. (2020). "Special report: The simulations driving the world's response to COVID-19". In: *Nature*. URL: https://www.nature.com/articles/d41586-020-01003-6 (visited on 10/15/2020).

Agyei, W., W. Obeng-Denteh, and E. A. Andaam (2015). "Modeling Nurse Scheduling Problem Using 0-1 Goal Programming: A Case Study Of Tafo Government Hospital, Kumasi-Ghana". In: *International Journal of Scientific & Technology Research* 4 (03).

Ahmed, M. A. and T. M. Alkhamis (2009). "Simulation optimization for an emergency department healthcare unit in Kuwait". In: *European Journal of Operational Research* 198 (3), pp. 936–942.

Alhazzani, W. et al. (2020). "Surviving Sepsis Campaign: guidelines on the management of critically ill adults with Coronavirus Disease 2019 (COVID-19)". In: *Intensive Care Medicine* 46 (5), pp. 854–887.

Arenas, M. et al. (2002). "Analysis Via Goal Programming of the Minimum Achievable Stay in Surgical Waiting Lists". In: *Source: The Journal of the Operational Research Society* 53 (4), pp. 387–396.

Ataollahi, F. et al. (2013). "A goal programming model for reallocation of hospitals' inpatient beds". In: *Middle East Journal of Scientific Research* 18 (11), pp. 1537–1543.

Azaiez, M. N. and S. S. Al Sharif (2005). "A 0-1 goal programming model for nurse scheduling". In: *Computers and Operations Research* 32 (3), pp. 491–507.

Bansal, S., B. T. Grenfell, and L. A. Meyers (2007). "When individual behaviour matters: homogeneous and network models in epidemiology". In: *Journal of the Royal Society Interface* 4.16, pp. 879–891.

Barasa, E. W., P. O. Ouma, and E. A. Okiro (2020). "Assessing the hospital surge capacity of the Kenyan health system in the face of the COVID-19 pandemic". In: *PLoS ONE* 15 (7 July).

Barthelemy, M. (2010). "Spatial Networks". In: *Physics Reports 499:1-101 (2011)*. arXiv: 1010.0302 [cond-mat.stat-mech].

Beigel, J. H. et al. (2020). "Remdesivir for the Treatment of Covid-19 — Final Report". In: *New England Journal of Medicine*.

Blake, J. T. and M. W. Carter (2001). *A goal programming approach to strategic resource allocation in acute care hospitals*.

Branas, C. et al. (2020). "Flattening the curve before it flattens us: hospital critical care capacity limits and mortality from novel coronavirus (SARS-CoV2) cases in US counties". Preprint.

Brand, S. et al. (2020). "Forecasting the scale of the COVID-19 epidemic in Kenya". Preprint.

Campbell, D., S. Marsh, and C. Bannock (2020). "London hospitals struggle to cope with coronavirus surge". In: *The Guardian*. URL: https://www.theguardian.

com/uk-news/2020/mar/20/london-hospitals-struggle-to-cope-with-coronavirus-surge (visited on 10/15/2020).

Carlitz, R. D. and M. N. Makhura (2021). "Life under lockdown: Illustrating tradeoffs in South Africa's response to COVID-19". In: *World Development* 137.

Chen, P-S. et al. (2015). *Scheduling Patients' Appointments: Allocation of Healthcare Service Using Simulation Optimization*.

Chow, D. and E. Saliba (2020). "Italy has a world-class health system. The coronavirus has pushed it to the breaking point." In: *NBC News*. URL: https://www.nbcnews.com/health/health-news/italy-has-world-class-health-system-coronavirus-has-pushed-it-n1162786 (visited on 10/15/2020).

Chu, D. K. et al. (2020). "Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis". In: *The Lancet* 395 (10242), pp. 1973–1987.

Data Science for Social Impact Research Group @ University of Pretoria (2020). *Coronavirus COVID-19 (2019-nCoV) Data Repository and Dashboard for South Africa*. URL: https://github.com/dsfsi/covid19za.

Davies, N. G. et al. (2020). "Age-dependent effects in the transmission and control of COVID-19 epidemics". In: *Nature Medicine* 26 (8), pp. 1205–1211.

Delobelle, P. (2013). "The health system in South Africa. Historical perspectives and current challenges". In: *South Africa in Focus: Economic, Political and Social Issues*.

Diop, B. Z. et al. (2020). "The relatively young and rural population may limit the spread and severity of COVID-19 in Africa: A modelling study". In: *BMJ Global Health* 5 (5).

Driessche, P. Van den (2017). "Reproduction numbers of infectious disease models". In: *Infectious Disease Modelling* 2.3, pp. 288–303.

Dzinamarira, T., M. Dzobo, and I. Chitungo (2020). "COVID-19: A perspective on Africa's capacity and response". In: *Journal of Medical Virology* 92 (11), pp. 2465–2472.

Feng, Y. Y., I. C. Wu, and T. L. Chen (2017). "Stochastic resource allocation in emergency departments with a multi-objective simulation optimization algorithm". In: *Health Care Management Science* 20 (1), pp. 55–75.

Ferland, J. A., B. Ahiod, and P. Michelon (2001). *Generalized Assignment Type Goal Programming Problem: Application to Nurse Scheduling*, pp. 391–413.

Gür, Ş. and T. Eren (2018). "Scheduling and planning in service systems with goal programming: Literature review". In: *Mathematics* 6 (11).

Haffajee, F. (2020). "Gauteng's oxygen shortages raise questions about lockdown planning". In: *Daily Maverick*. URL: https://www.dailymaverick.co.za/article/2020-07-09-gautengs-oxygen-shortages-raise-questions-about-lockdown-planning/ (visited on 09/28/2020).

Harding, A. (2020). "Coronavirus in South Africa: Inside Port Elizabeth's 'hospitals of horrors'". In: *BBC*. URL: https://www.bbc.com/news/world-africa-53396057 (visited on 09/28/2020).

He, X. et al. (2020). "Temporal dynamics in viral shedding and transmissibility of COVID-19". In: *Nature Medicine* 26 (5), pp. 672–675.

Hens, N. et al. (2010). "Seventy-five years of estimating the force of infection from current status data". In: *Epidemiology and Infection* 138 (6), pp. 802–812.

Horby, P. et al. (2020). "Effect of Dexamethasone in Hospitalized Patients with COVID-19: Preliminary Report". In: *The New England Journal of Medicine*.

Hu, H., K. Nigmatulina, and P. Eckhoff (2013). "The scaling of contact rates with population density for the infectious disease models". In: 244, pp. 125–134.

Huarng (1999). "Computers in Nursing A Primary Shift Rotation Nurse Scheduling Using Zero-One Linear Goal Programming". In: 17 (3), pp. 135–144.

Ismail, A. (2020). "Covid-19 reinfection cases in SA: NICD can't confirm repeat infections, wants labs to keep specimens". In: *News24*. URL: https://www.news24.com/news24/southafrica/news/covid-19-reinfection-cases-in-sa-nicd-cant-confirm-repeat-infections-wants-labs-to-keep-specimens-20201117 (visited on 01/31/2021).

Iwasaki, A. (2020). "What reinfections mean for COVID-19". In: *The Lancet Infectious Diseases*.

Jenal, R. et al. (2011). "A cyclical nurse schedule using goal programming". In: *Journal of Mathematical and Fundamental Sciences* 43A (3), pp. 151–164.

Jerbi, B. and H. Kamoun (2009). "Using simulation and goal programming to reschedule emergency department doctors' shifts: Case of a Tunisian hospital". In: *Journal of Simulation* 3 (4), pp. 211–219.

Johns Hopkins University and Medicine (2021). *COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)*. URL: https://coronavirus.jhu.edu/map.html (visited on 03/13/2021).

Kermack, W. O. and A. G. McKendrick (1927). "A Contribution to the Mathematical Theory of Epidemics". In: *Proceedings of the Royal Society of London Series A* 115.772, pp. 700–721.

Kommenda, N. and F. Hulley-Jones (2020). "Covid vaccine tracker: when will a coronavirus vaccine be ready?" In: *The Guardian*. URL: https://www.theguardian.com/world/ng-interactive/2020/oct/20/covid-vaccine-tracker-when-will-a-coronavirus-vaccine-be-ready (visited on 10/27/2020).

Kucharski, A. J. et al. (2020). "Early dynamics of transmission and control of COVID-19: a mathematical modelling study". In: *The Lancet Infectious Diseases* 20 (5), pp. 553–558.

Kwak, N. K. and C. Lee (1997). *A Linear Goal Programming Model for Human Resource Allocation in a Health-Care Organization*.

Lal, T. M., T. Roh, and T. Huschka (2016). "Simulation based optimization: Applications in healthcare". In: vol. 2016-February. Institute of Electrical and Electronics Engineers Inc., pp. 1261–1271.

Lepule, T. (2020). "Western Cape hospitals buckling under pressure of rising staff shortages". In: *IOL*. URL: https://www.iol.co.za/weekend-argus/news/western-cape-hospitals-buckling-under-pressure-of-rising-staff-shortages-50387769 (visited on 10/15/2020).

Leung, K. et al. (2020). "First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment". In: *The Lancet* 395 (10233), pp. 1382–1393.

Long, E.F., E. Nohdurft, and S. Spinler (2018). "Spatial Resource Allocation for Emerging Epidemics: A Comparison of Greedy, Myopic, and Dynamic Policies". In: *Manufacturing and Service Operations Management*.

Low, M. and N. Geffen (2020). "8 different models predict how coronavirus cases could increase in South Africa". In: *BusinessTech*. URL: https://businesstech.co.za/news/trending/400597/8-different-models-predict-how-coronavirus-cases-could-increase-in-south-africa/ (visited on 09/24/2021).

Lucidi, S. et al. (2016). "A Simulation-Based Multiobjective Optimization Approach for Health Care Service Management". In: *IEEE Transactions on Automation Science and Engineering* 13 (4), pp. 1480–1491.

Mahlathi, P. and J. Dlamini (2015). *Minimum data sets for human resources for health and the surgical workforce in South Africa's health system*.

Malik, T. and O. Sharomi (2017). "Optimal control in epidemiology". In: *Annals of Operations Research*.

Mbah, M.L.N. and C.A. Gilligan (2011). "Resource Allocation for Epidemic Control in Metapopulations". In: *PLoS ONE*.

Mkhize, Z. (2020). *Latest confirmed cases of COVID-19 in South Africa (31 Oct 2020)*. URL: https://www.nicd.ac.za/latest-confirmed-cases-of-covid-19-in-south-africa-31-oct-2020/ (visited on 03/13/2021).

Moon, Y. et al. (1989). *Application of goal programming to improve resource allocation for health services in Papua New Guinea*, pp. 81–95.

Mukandavire, Z. et al. (2020). "Quantifying early COVID-19 outbreak transmission in South Africa and exploring vaccine efficacy scenarios". In: *PLoS ONE* 15 (7 July).

Mushayabasa, S., E. T. Ngarakana-Gwasira, and J. Mushanyu (2020). "On the role of governmental action and individual reaction on COVID-19 dynamics in South Africa: A mathematical modelling study". In: *Informatics in Medicine Unlocked* 20.

National Department of Health (2020). "EMS Readiness Plans for COVID-19 Surge".

Nyabadza, F. et al. (2020). "Modelling the potential impact of social distancing on the COVID-19 epidemic in South Africa". In: *Computational and Mathematical Methods in Medicine*.

Oddoye, J. P. et al. (2009). "Combining simulation and goal programming for healthcare planning in a medical assessment unit". In: *European Journal of Operational Research* 193 (1), pp. 250–261.

Oliveira, B. R.P. e et al. (2020). "A Simulation-Optimisation approach for hospital beds allocation". In: *International Journal of Medical Informatics* 141.

Ozcan, Y. A., E. Tànfani, and A. Testi (2017). "Improving the performance of surgery-based clinical pathways: a simulation-optimization approach". In: *Health Care Management Science* 20 (1).

Pei, S. and J. Shaman (2020). "Initial Simulation of SARS-CoV2 Spread and Intervention Effects in the Continental US". Preprint.

Peng, L. et al. (2020). "Epidemic analysis of COVID-19 in China by dynamical modeling". Preprint.

Petersen, T. (2020). "25 ICU beds for Covid-19 patients 'already full' in Cape Town's largest hospital". In: *News24*. URL: https://www.news24.com/news24/southafrica/news/icu-beds-for-covid-19-patients-already-full-in-cape-towns-largest-hospital-20200521 (visited on 09/28/2020).

Rauner, M. S. and N. Bajmoczy (2003). "How many AEDs in which region? An economic decision model for the Austrian red cross". In: vol. 150, pp. 3–18.

Ripperger, T. J. et al. (2020). "Orthogonal SARS-CoV-2 Serological Assays Enable Surveillance of Low Prevalence Communities and Reveal Durable Humoral Immunity." In: *Immunity*.

RSA Government (2008). "Ambulance Services 2009". In: *Government Gazette*.

– (2020a). "Regulations issued in terms of Section 27(2) of the Disaster Management Act, 2002". In: *Government Gazette* Vol. 657 (No. 43107).

– (2020b). "Disaster Management Act 2002: Amendment of Regulations issues in terms of Section 27(2)". In: *Government Gazette* Vol. 657 (No. 43148).

– (2020c). "Disaster Management Act 2002: Amendment of Regulations issues in terms of Section 27(2)". In: *Government Gazette* (No. 43232).

– (2020e). "Disaster Management Act 2002: Amendment of Regulations issues in terms of Section 27(2)". In: *Government Gazette* Vol. 658 (No. 43258).

– (2020f). "Disaster Management Act 2002: Amendment of Regulations issues in terms of Section 27(2)". In: *Government Gazette* (No. 43364).

– (2020g). "Disaster Management Act 2002: Amendment of Regulations issues in terms of Section 27(2)". In: *Government Gazette* (No. 43620).

– (2020h). "Disaster management act, 2002: Amendment of regulations issued in terms of Section 27(2)". In: *Government Gazette* (No. 43725).

SA COVID-19 Modelling Consortium (2020a). "National COVID Epi Model". In: URL: https://sacovid19mc.github.io.

– (2020b). *Estimating cases for COVID-19 in South Africa: Assessment of alternative scenarios*. URL: https://www.nicd.ac.za/wp-content/uploads/2020/11/SACovidModellingReport_LongTermProjections_050920_final.pdf.

SA Provincial Health (2020). "Eastern Cape hospitals crisis: 'Patients fight one another for oxygen'". In: *Medical Brief*. URL: https://www.medicalbrief.co.za/archives/eastern-cape-hospitals-crisis-patients-fight-one-another-for-oxygen/ (visited on 10/15/2020).

Sanyaolu, A. et al. (2020). "Comorbidity and its Impact on Patients with COVID-19". In: *SN Comprehensive Clinical Medicine* 2 (8), pp. 1069–1076.

Serrato-Garcia, M. A., J. Mora-Vargas, and R. T. Murillo (2016). "Multi objective optimization for humanitarian logistics operations through the use of mobile technologies". In: *Journal of Humanitarian Logistics and Supply Chain Management* 6 (3), pp. 399–418.

Soetaert, K., T. Petzoldt, and R. W. Setzer (2010). "Solving Differential Equations in R: Package deSolve". In: *Journal of Statistical Software* 33.9, pp. 1–25. URL: http://www.jstatsoft.org/v33/i09.

Statistics South Africa (2018). *General Household Survey*.

– (2020). *Mid-year population estimates*. URL: www.statssa.gov.za,info@statssa.gov.za,Tel+27123108911.

Stewart, T. J. (2007). *Principles and Practice of Linear Programming*.

Taboe, H. B. et al. (2020). "Predicting COVID-19 spread in the face of control measures in West Africa". In: *Mathematical Biosciences* 328.

The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team (2020). *The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China*.

Theussl, S. and K. Hornik (2019). *Rglpk: R/GNU Linear Programming Kit Interface*. R package version 0.6-4. URL: https://CRAN.R-project.org/package=Rglpk.

Tillett, R. L. et al. (2020). "Genomic evidence for reinfection with SARS-CoV-2: a case study". In: *The Lancet Infectious Diseases*.

Van Den Heever, A. (2020). Social Security Systems Administration and Management Studies, University of Witwatersrand, Johannesburg. Unpublished data.

Wiersinga, W. J. et al. (2020). "Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review". In: *JAMA - Journal of the American Medical Association* 324 (8), pp. 782–793.

Zhao, Z. et al. (2020). "Prediction of the COVID-19 spread in African countries and implications for prevention and control: A case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya". In: *Science of the Total Environment* 729.