Robust facility location of container clinics: A South African application

Carike Karsten

A report in fulfilment of the requirements for the degree

MASTER OF ENGINEERING (INDUSTRIAL ENGINEERING)

in the

FACULTY OF ENGINEERING, BUILT ENVIRONMENT AND INFORMATION TECHNOLOGY

UNIVERSITY OF PRETORIA

July 22, 2021

DECLARATION OF ORIGINALITY

Full names: Carike Karsten
Student number: 15012396
Declaration:

- I understand what plagiarism is and I am aware of the University‘s policy in this regard.
- I declare that this is my own original work.
- Where other people‘s work has been used (either from a printed source, internet, or any other source) this has been carefully acknowledged and referenced in accordance with departmental requirements.
- I have not used another student‘s past work to hand in as my own.
- I have not allowed and will not allow anyone to copy my work with the intention of handing it in as their own work.

Signature: Carike Karsten
(By typing my name, I pledge that I signed this declaration.)

Executive summary

Health care, and especially access to health care, has always been a critical metric for countries. In 2017, South Africa spent 9% of its GDP on health care. Despite the GDP health care allocation being 5% higher than recommended by the World Health Organisation for a country of its socio-economic status, South Africa's health status is poor compared to similar countries. In 1994, South Africa implemented a health care policy to make health care accessible to all South Africans. A primary health care facility within 5 km of the place of residence is deemed accessible. There is still a significant gap between the actual and desired accessibility, especially for the lower-income communities. There is a need to improve access to public health care for all South Africans. Cost-effective and sustainable solutions are required to solve this problem. Therefore, an opportunity was identified to investigate the location of low-cost container clinics in lower-income communities.

This report uses robust optimisation and goal programming to find robust sites for cost-effective container clinics over multiple years in an uncertain environment using multiple future city development scenarios. The study area of the report includes three metro municipalities (City of Tshwane, City of Johannesburg, and City of Ekurhuleni) in Gauteng, South Africa. Three future development scenarios were created for this study using a synthetic population and urban growth simulation model developed by the Council for Scientific and Industrial Research (CSIR). The model provided the population distribution from 2018 to 2030 for all three of the scenarios.

The simulation model provides household attribute tables as an output. Household attributes that have a causal relationship with health care demand were investigated during the literature review. Based on the literature and the available household attributes, four attributes were selected to forecast the health care demand. The four attributes are household income, the number of children in the household, the household size, and the nearest clinic's distance.

Using associative forecasting, the primary health care demand was forecasted from 2018 to 2030. These forecasts were used as input into the facility location models. A p-median facility location model was developed and implemented in Python. Since facility location problems are classified as NP-hard problems, heuristics and metaheuristics were investigated to speed up the problem solving. A Genetic Algorithm (GA) selected as the metaheuristic be used to determine a suitable configuration of facilities for each scenario. The model determined good locations of clinics from a set of candidate locations. A good year to open each clinic is also determined by the model. These decisions are made by minimising three variables: total distances travelled by the households to their nearest clinics, the total distance from the selected distribution centre to the open clinics and the total building cost. An accessibility target of 90% was added to the model to ensure that at least 90% of the households are within 5 km of the nearest clinic within the first five years. In these models, operating costs were not included. Therefore all the results are skewed, with most of the clinics being opened in the first year when it is the cheapest since there is no penalty for opening a clinic before it is needed — the exclusion of operating costs is a shortcoming to address in future work.

A goal programming model was developed with the variables of the individual scenarios as the goals. The goal programming model was implemented in Python and used to determine a robust configuration of where and in what year to open container clinics. A difference of 25% was set as the upper limit for the difference between the robust configuration variables and the good or acceptable variables for the individual scenarios as the scenarios investigated are very different. This ensured that the robust solution would perform well for any of the three scenarios. The model was able to find locations that provided a relatively good solution to all the scenarios. This came with a cost increase, but that is a trade-off that must be made when dealing with uncertainty. This model is a proof of concept to bridge the gap between urban planning with multiple development scenarios and facility location, more specifically robust facility location.

The biggest rendement was achieved by constructing and placing the container clinics in the shortest space of time because the 90% accessibility requirement can be addressed cost-effectively without an operating cost penalty — this is unfortunately not possible in reality due to budget constraints. An accessibility analysis was conducted to investigate the impact of the accessibility percentage on the variable values and to test the model in a scenario closer resembling the real world by adding a budget constraint. The time limit of the accessibility requirement was removed. In this case, a gradual improvement in the accessibility over the 12 years was observed due to the gradual opening of clinics over the years. Based on the analyses results, it was concluded that the model is sensitive to changes in parameters and that the model can be used for different scenarios.

Acknowledgements

It has truly been an incredible journey of learning and developing. It would not have been possible on my own, and I would like to express my sincere gratitude towards the following role players:

- My heavenly Father for the ability and perseverance throughout the process
- My mother and father for their guidance and support throughout the process
- My sisters for their motivation and continual support
- Mr Quintin van Heerden at the Council for Scientific and Industrial Research for his mentorship, support, and belief in me
- Dr Wilna Bean, my supervisor at the Department of Industrial and Systems Engineering at the University of Pretoria, for her guidance throughout the completion of my dissertation

Contents

List of Figures

List of Tables

List of Algorithms

Acronyms

BEPP	Built Environment Performance Plan
CSIR	Council for Scientific and Industrial Research
EMM	Ekurhuleni Metropolitan Municipality
GA	Genetic Algorithm
GDP	Gross Domestic Product
IDP	Integrated Development Plan
ITP	Integrated Transport Plan
MSDF	Metropolitan Spatial Development Framework
SA	Simulated Annealing
TS	Tabu Search
WHO	World Health Organisation

Chapter 1

Introduction

Health care, and especially access to health care, has always been a key metric for countries. In 2017, 9% of South Africa's Gross Domestic Product (GDP) was spent on health care. Despite this being 5% higher than recommended by the World Health Organisation (WHO) for a country of its socio-economic status, the country's health is poor compared to similar countries. Two of the reasons for this imbalance are inequalities between the public and private health care systems, and restricted access to health care in some communities (Africa Health, 2019).

South Africa has a two-tiered health care system. The first tier is the large public sector used by the majority of citizens. These citizens are not members of medical aid schemes. According to Stats SA (2018), 83.6% of South Africans are not part of a medical aid scheme. The second tier is the growing private sector, utilised mainly by members of medical aid schemes (Africa Health, 2019). Due to the high costs of private health care, a large percentage of the population is forced to use public health care. Over 80% of the lowest income quantile make use of public health care facilities when ill, while more than 70% of the highest income quantile make use of private health care facilities (Booysen, 2003).

According to the South African Constitution, access to health care is a constitutional right for all South Africans (South Africa, 1996). Therefore, the government needs to ensure accessibility to decent health care for all South Africans. There are significant differences in the various households' proximity to a health care facility between rural and urban areas as well as socio-economic groups (Booysen, 2003). Many people using public health care are located far from hospitals or clinics, even within more urbanised areas. According to a survey done by Yantzi et al. (2001), distance to hospitals and clinics is a crucial factor when selecting health services and whether to visit. The convenience of access to a health care facility is a critical factor in deciding whether or not to visit a health care facility. To make health care more accessible, mobile clinics are often proposed. Mobile clinics are customised vehicles that can travel into communities to provide immediate but transient health care to the people in these communities (Hill et al., 2014).

Another alternative for providing accessible health care is container clinics, i.e. shipping containers converted into clinics. This is a more permanent alternative but without the extensive cost implications. These container clinics adhere to the National Building Regulations (IUSS, 2014). Investing in container clinics rather than mobile clinics can provide a sense of security; the community knows that it will not disappear overnight and can be accessed regularly.

When an investment is made into container clinics, it should be noted that efficient facility location could improve the utilisation of the facilities (Peng and Afshari, 2014). Therefore, locating clinics on commuter routes are advantageous.

Once the facilities have been built, medicine is required to provide the patients with the needed health care. Delivering medicine to a clinic is usually at the expense of the pharmaceutical company. The expertise of pharmaceutical companies relates to improving their product safety and quality. Logistics and distribution expertise are often lacking in these companies (Ahmadi-Javid et al., 2017) and associated government departments. Distribution network costs are continuously increasing and consume a significant amount of a company's budget. Warehousing and transportation costs comprise 15.2%, and 57% of logistics costs respectively (Havenga et al., 2016). Reducing these costs is in the interest of the patients since these costs could ultimately be passed down to them through the product cost (Ahmadi-Javid et al., 2017).

Considering these large and increasing distribution network costs, the need for urban distribution network planning is more significant than ever. Considering distribution network planning and urban growth simultaneously can assist in better strategic decision making for a company or government department. Distribution network robustness is a crucial strategic consideration for all companies (Graham et al., 2015). A robust distribution network's configuration ensures that the performance level or accessibility level of the network stays at the desired level irrespective of changes in the customer base, often due to changes in population growth rates or urbanisation.

The population growth of South Africa is roughly 1.4% per year (Macrotrends, 2020), which leads to densification and urban growth. Urban growth is a result of population growth in urban areas as well as the expansion of cities in the surrounding areas. This growth results from general population growth and people relocating from rural areas to urban areas, mainly for improved work opportunities. Urban growth directly influences a city's liveability, economic opportunities, and sustainability, as well as the demand for goods and services such as food, pharmaceuticals, and medical services (Tank, 2017).

This report investigates robust clinic placement for container clinics in low-income and lower-middle-income communities to improve access to primary health care and reduce total building cost, travelling and distribution distances. According to the 2011 South African census, low-income households have an annual income between R 1 and R 19 200. Middle-income households have an annual income of R 19 201 to R 307 200 (Stats SA, 2011). The annual household income threshold for a lower-middle-income household is R 108 000. The households in these income categories were selected for this study as it is these households that will most likely make use of public health care. Container clinics are used in public health care to improve access to health care; therefore, they were selected as the facility type used for this study.

The study area for this report includes three metros in Gauteng, South Africa. These three metros are the City of Tshwane, the City of Johannesburg and the City of Ekurhuleni in Gauteng, South Africa, as shown in Figure 1.1. These metros were selected for the study area since the data required for the investigation was readily available for these three metros. Gauteng is home to just more than a quarter of the South African population (Stats SA, 2020). The three selected metros have the highest population density in the province. Therefore, the proportion of the population can be maximised while keeping the case study areas relatively small.

All municipalities are required to do planning in the form of a master plan that includes the Integrated Development Plan (IDP) and the Metropolitan Spatial Development Framework (MSDF). This plan is a long-term dynamic planning document providing a conceptual layout to guide future municipal development and growth (The World Bank Group, 2015). Possible revenues, expenses and development projects are a few of the topics included in these plans. Ekurhuleni Metropolitan Municipality (EMM) has an Aerotropolis master plan where all their blueprints to become an aerotropolis (economic hub around the O.R. Tambo International Airport) are laid out. In this plan, there are 21 development projects in priority areas to ensure growth and stimulate the economy in the next 20 years (Ekurhuleni Metropolitan Municipality, 2015). These development projects are either housing or job creation projects. The deployment sequence of these projects can affect the growth of the municipality. The other two municipalities have similar documents, however, not in the same detail.

Although the municipalities have these development plans, not all projects realise for numerous reasons, such as budget constraints and shifts in focus or importance. Development project deployment strategies are combinations of implemented development projects. The realisation of deployment strategies ultimately determines the development patterns of the municipality and the city form.

Urban growth simulation modelling is used as a tool to test the likely outcomes of development project deployment strategies. These models show possible future city form scenarios based on the developments as well as other models such as transportation and location choice models. A city's form refers to the physical characteristics of a city, including spatial distribution of households, work opportunities and other land use. To this extent, the Council for Scientific and Industrial Research (CSIR) has implemented an urban growth simulation model (UrbanSim) based on synthetic population data for the three metros of Gauteng.

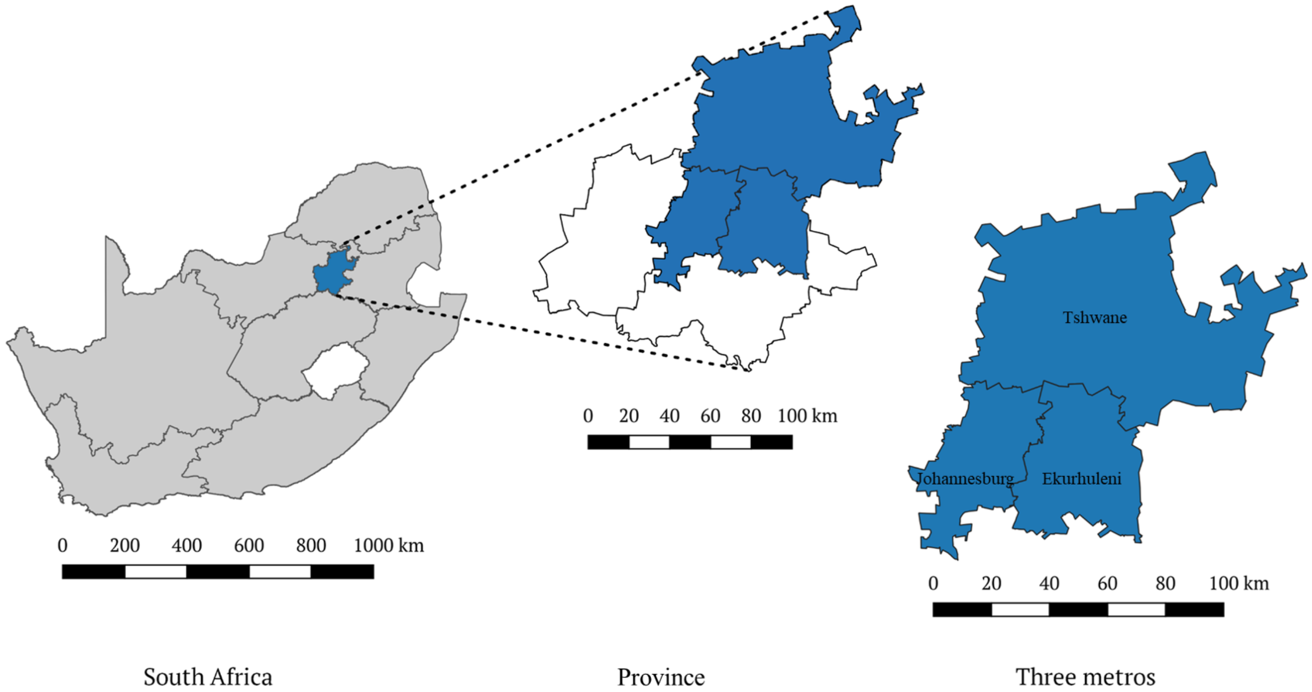


Figure 1.1: Three Gauteng metros used in the analysis

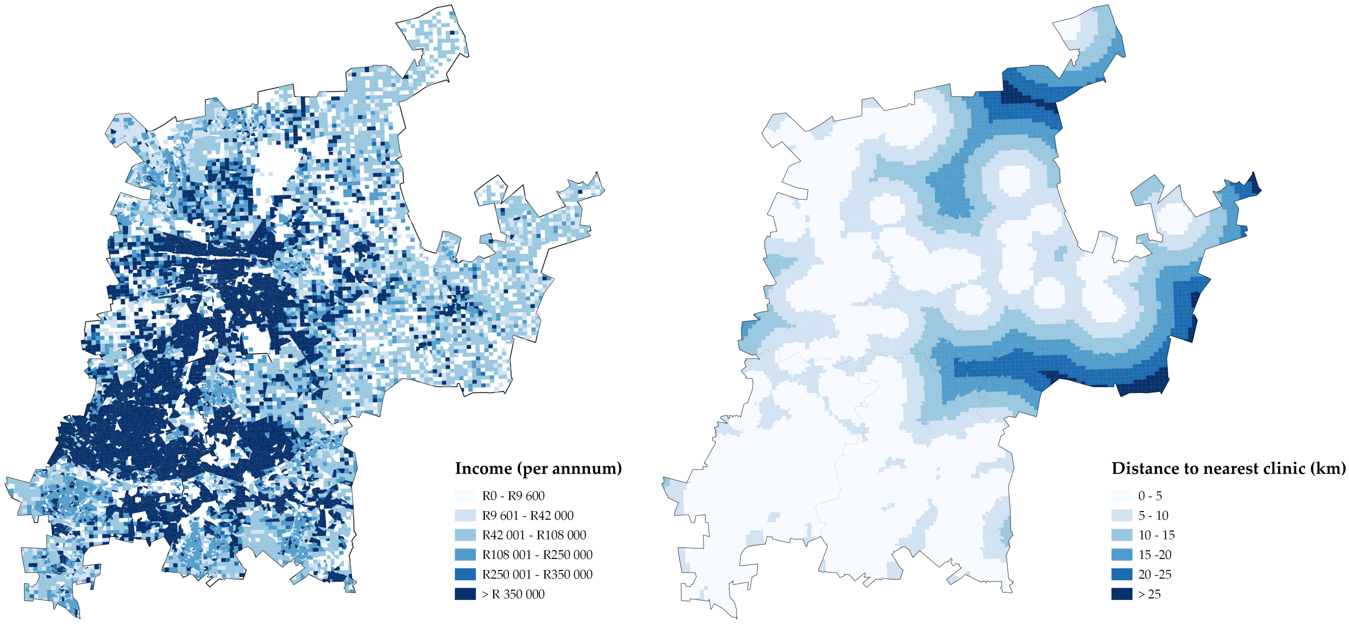
Three scenarios were developed and tested for the three metros with various development project alternatives and schedules. The model is used to make predictions from 2019 to 2030, based on the validation done on data from 2011 to 2018. The possible future city forms and results from the simulations were made available to work from in this project. These simulation results consist of numerous tables of which one is the simulated synthetic household table. This table contains the household's location and attributes that will be used to derive the demand for primary health care.

1.1 Research opportunity

South Africa implemented two health care policies in 1994. The first policy is free health care for pregnant women and children under six years of age. The second is accessible primary health care for all South Africans. In 2001, the South African Department of Health developed a comprehensive primary health care service package. This package aims to provide all South Africans with primary health care no more than 5 km from their place of residence (Nteta et al., 2010). South Africans using public health care have restricted access to health care facilities due to the distribution of the facilities across the country and public transportation challenges. This service plan has improved accessibility to health care facilities, however, there is still a significant gap between the actual and desired accessibility for the lower-income communities. There is a need to improve access to public health care for all South Africans. Public health care is available to all citizens and is subsidised by the government. In contrast, private health care is only available to a small percentage of the population who have the means to pay. Lower-income households are more likely to make use of public health care since they do not have the means to pay for private health care (Young, 2016). Cost-effective and sustainable solutions are required to solve this problem. Therefore, an opportunity was identified to investigate the location of container clinics within walking distance to lower-income communities that are more likely to use public health care than private health care.

The current distribution of household income per annum and accessibility to health care facilities in the three metros are depicted in Figure 1.2. The income distribution of the three metros can be seen in Figure 1.2a. In Figure 1.2b, the distance to the nearest public health care facility is illustrated.

By comparing these two figures side by side, the areas with the furthest distances to public health care facilities are generally in lower-income areas. There is an exception in Johannesburg: In this metro, most households, regardless of income, are within 5 km of a public health care facility. In contrast to this, there is still great inequality in access to health care based on income. Ekurhuleni also has some inequality, however, not nearly as much as Tshwane.



(a) Household income distribution across the three metros (b) Distance to nearest health care facility across the metros

Figure 1.2: Heatmaps of income distribution and distance to the nearest clinics for the three metros

Accessibility to public health care needs to be improved, especially for lower-income communities. Using container clinics is a potential cost-effective solution. This accessibility improvement must happen in a financially responsible manner and, therefore, a trade-off between cost and accessibility must be made during the decision to locate container clinics.

From the research opportunity, the research question was formulated: *Where and when should the container clinics be placed given multiple possible future scenarios to minimise the total travel distance for patients, the distribution companies and the building cost?*

1.2 Research design

The main objective of this report is to find robust locations for container clinics in lower- to medium-income and lower-income communities and determine the years in which they should be opened based on various future development scenarios. For this health care demand, forecasts for the years in question are required. Based on the knowledge gained from the literature study, associative forecasting can be used to forecast the primary health care demand based on available household attributes for three scenarios. This demand forecast is done per scenario and will be used as input into two models developed in this project. The first model is a scenario optimisation model. In this model, the demand forecast per scenario can now be used as input to determine a good configuration of the clinics by minimising the variable values for that scenario. The second model is the robust facility location model. This model use demand forecast per scenario and the minimised variable values of the scenario optimisation model as inputs to minimise the difference between the robust solution and the scenario solutions. The robust model determines a configuration of clinics that will perform well in all three scenarios.

The methodology for this research design is discussed in the next section.

1.3 Research approach

This project is defined as a multiple robust facility location project, focusing on finding robust locations for container clinics for multiple possible future scenarios for the three metros. The word 'robust' refers to a solution that will perform well given any possible realisation of random parameters. The steps that will be followed to ensure success are listed below. A detailed methodology is provided in Chapter 3 of this report.

1. **Opportunity analysis:** The research opportunity is identified in Section 1.1. There is a need to identify robust locations for container clinics in lower-income communities. For the development of a robust solution, various future scenarios have to be considered.

2. **Methodology:** Literature is investigated in Chapter 2, to determine which household attributes can be used to predict health care demand. In Chapter 3, the information gained from the literature study is used to convert the available and relevant household attributes into health care demand. The demand is used as input for the robust facility location model to determine the most robust locations for the container clinics.

3. **Model development:** A multiple facility location algorithm is developed to determine the good locations for each scenario. Thereafter, a goal programming algorithm is developed to determine the most robust locations. The model development is described in Chapter 4.

4. **Evaluation:** In this step, verification and validation of the model and solution is done. The evaluation of the solution is discussed in Chapter 5.

5. **Accessibility and budget analyses:** Some of the constraints are altered in Chapter 6 to determine if the model is useful for various cases.

1.4 Limitations

The scope of the project is inhibited by data availability. Only metros for which the synthetic population is available could be used. The UrbanSim model outputs constrain this project. Numerous factors affect health care demand, however, only the household attributes provided by the simulation model are considered in this project.

1.5 Expected contribution

This report aims to fill the gap in the literature between facility location, especially robust facility location and urban planning. A proof of concept is developed, in the form of a python model, to determine robust locations, specifically for container clinics, when urban planners and key role players consider multiple future development strategies.

1.6 Document structure

In Chapter 2, a critical literature review is conducted on the correlation between household factors and health care demand. Health care facility location models and robust solution alternatives are also investigated. Chapter 3 stipulates the process followed to forecast health care demand using household attributes as explanatory variables. In this chapter, the methodology to determine robust locations is stipulated. The model is developed in Chapter 4. In Chapter 5, the solutions for each of the scenarios as well as the robust solution are determined and investigated. The results of the accessibility and budget analyses of the model are reflected in Chapter 6. Lastly, the report is concluded in Chapter 7, and future work is described.

Chapter 2

Literature review

This chapter contains a literature review of household factors that can be used in associative forecasting of health care demand. After that, facility location models with and without uncertainty are explored. Methods to deal with uncertainty and especially robust models and solution methods are investigated. Finally, heuristics are researched to assist in solving the facility location model.

2.1 Health care demand forecasting

Forecasting can be done using one of two methods: qualitative forecasting and quantitative forecasting. Quantitative forecasting relies on data, while qualitative forecasting relies more on estimates and opinions. The quantitative method can be broken down into two categories: historical forecasts and associative forecasts. Historical forecasts make use of historical data and trends to predict future demand. A study by Soebiyanto et al. (2010) made use of a time series model, a historical forecasting technique, to predict seasonal influenza transmissions. Jones et al. (2008) also made use of the time series model to forecast daily patient volumes in a hospital emergency department. It was concluded that with enough data, historical forecasting methods and especially time series forecasting provides relatively accurate forecasts for demand with seasonal or weekly patterns. Associative forecasts identify causal relationships between variables (Jain, 2005). Actual demand data was not available to work from for historical forecasts. However, synthetic household attribute data was made available and, therefore, associative forecasting was selected to determine the health care demand. In this section, household attributes with a causal relationship with health care demand are investigated.

2.1.1 Factors influencing health care utilisation

It is believed that the demand for health care is closely related to the health-seeking behaviour of individuals (Sarma, 2009). There are many socio-economic characteristics of individuals that affect their health-seeking behaviour (Nahu, 2006) and some observable socio-economic characteristics were investigated to gain an understanding of how these factors can affect health care demand.

Availability and affordability are two of the critical decision influencers when it comes to seeking health care, especially in developing countries. The effect of cost on health care demand is much higher in developing countries because a large proportion of the population has a lower income and no medical aid (O'donnell, 2007).

Health care affordability is primarily influenced by two factors: household income and health care cost. In a study done by Mwabu et al. (1993), a strong positive correlation between income and the probability of seeking medical care compared to self-treatment was found. People in the upper and middle socio-economic classes are more likely to seek medical help than people in the lower socio-economic class (Abera Abaerei et al., 2017).

Other factors related to income and affordability are investigated in the literature. Higher-income classes are more likely to be part of a medical aid scheme. People who are part of a medical aid scheme are more likely to seek medical care from a health care facility than people who are not (Abera Abaerei et al., 2017).

Even when free health care is provided, the monetary and time cost of travel to the local clinic is seen as a health care cost (McLaren et al., 2014). Travelling cost is also identified as a reason why health care facilities are not visited. In some communities, residents do not have the finances to pay the travelling costs to far away clinics as well as the out-of-pocket payments at the facility (Nteta et al., 2010).

In studies done by Mwabu et al. (1993) and Hotchkiss (1998), a negative correlation was found between seeking medical care and the distance to the medical care facility. A study conducted in KwaZulu-Natal, South Africa, found that households within 30 minutes from a clinic are ten times more likely to visit the facility when ill than those further away (Booysen, 2003). The main contributors to this distance decay are gender, age, income, access to transportation and other socio-economic factors. (Buor, 2003; Nemet and Bailey, 2000). Mitropoulos et al. (2013) and Verter and Lapierre (2002) investigated the location of preventative health care facilities using an exponential distance decay function to determine the probability of patients visiting a clinic. The distance decay used in these studies can be seen in the graph in Figure 2.1. From this distance decay graph, it can be deduced that distance indeed significantly impacts the probability of visiting a primary health care facility.

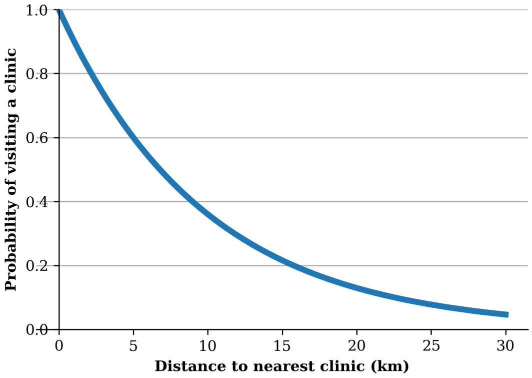


Figure 2.1: Distance decay function used by Mitropoulos et al. (2013) and Verter and Lapierre (2002)

Tanser et al. (2006) and Cooke et al. (2010) found that households further than 5 km from the health care facility are half as likely to utilise the health care provided, even for life-saving anti-retroviral therapy for HIV/AIDS. Therefore, if this is the case for life-saving treatment, the same or even worse statistics can be expected to treat less life-threatening diseases or conditions.

Another factor identified in the literature is the age of the patient. According to Masiye and Kaonga (2016), there is a slight negative correlation between age and the likelihood of seeking proper medical care rather than self-medicating. The study done by Abera Abaerei et al. (2017) found that a one-year increase in age increases the odds of seeking medical care by 2%. Similar results were found in a study done by Nteta et al. (2010); they found a positive correlation between age and the utilisation of health care facilities in Tshwane municipality in Gauteng, South Africa.

In 1994, a health care policy was implemented in South Africa to provide free health care for mothers and children under the age of six years (Nteta et al., 2010). A study done by Abera Abaerei et al. (2017), found that females are almost as likely as men to visit a health care facility.

When considering maternal health care visits, a study done by Wabiri et al. (2016) found that the probability of a pregnant woman visiting the doctor four or more times while pregnant increases with age and an increase in education level. In this study, the pregnant woman's employment status also impacted the frequency of check-ups. It was more likely for employed women to go for regular check-ups at a clinic or a doctor than unemployed pregnant women.

Wellay et al. (2018) concluded that the perceived quality of health care provided was a statistically significant variable when determining is a household will seek medical advice when ill. Abera Abaerei et al. (2017) found that if a patient was satisfied with the quality of care provided, they were more likely to seek medical help again in the future. Health care quality is often a subjective opinion and not necessarily a quantifiable household attribute that can be used.

Various attributes contribute to health care utilisation. Most of these factors investigated are household attributes. The highest correlation factors are distance to the facility and affordability. The other factors such as age, gender, education, and employment status also have an impact, however, it is less significant. With the household attributes that have the most significant impact on health care demand having been identified, facility location models are investigated to serve the demand.

2.2 Facility location models

Facility location is an operations research branch concerned with locating at least one facility to optimise (maximise or minimise) at least one objective function such as cost, coverage, or travel distance (Farahani et al., 2010). Facility location problems are generally classified as continuous or discrete. Continuous implies that facilities may be placed anywhere in the feasible region. Discrete implies that facilities can only be located at candidate locations (Ahmadi-Javid et al., 2017). In this study, the focus will be on discrete facility location models since these models are most often used in the health care industry (Meskarian et al., 2017) and there is a fixed set of sites where the facilities can be placed.

Discrete location modelling assumes that individual demand points can be grouped into selected discrete demand points. This enables the modeller to represent a geographical area with a few hundred demand points, rather than thousands of demand points. Another assumption made when using discrete location models is that there is a finite set of candidate locations where the facilities can be placed (Meskarian et al., 2017).

In discrete location modelling, many models can be applied in the health care facilities industry. Discrete location modelling is divided into three categories: covering-based problems, median-based problems and other (Meskarian et al., 2017).

Covering-based models assume a critical coverage time or distance within which the demand must be served to be considered covered. This class of discrete location models includes the p-centre model, the maximal-covering model, and the set-covering model (Daskin, 2008). These models are min-max type of problems and are often referred to as location-allocation problems since the facility location and demand allocation are done simultaneously. In the health care environment, covering-based models are most often used for locating emergency service facilities (Ahmadi-Javid et al., 2017).

The set covering model was defined by ReVelle et al. (1976) and aims to minimise the number of sites needed or total cost to cover all the demand points. However, in this model there is a limited set of facilities that can be selected. This limitation can make it challenging to achieve full coverage (Peng and Afshari, 2014). The covering model aims to maximise the number of demand points covered with a given set of sites (Daskin, 2008). A basic set cover model will minimise the cost of the facilities needed to cover the demand point, with a constraint to ensure that all demand points are covered (Ahmadi-Javid et al., 2017).

The maximal covering model aims to maximise coverage with a restrictive number of facilities that can be located within a predefined maximum coverage distance (Church and ReVelle, 1974). This model is popular in developing countries and rural areas since households are more widespread and it is not financially feasible to locate a facility close to every household (Rahman and Smith, 2000). Marianov and Taborga (2001) applied this maximal covering model in their study and the model located public health care facilities to maximise the coverage of lower-income households, while still catering for the high-income population. Taiwo (2020) to identify optimal or near optimal locations for COVID-19 in Nigeria.

The p-centre model aims to minimise the maximum distance to the facility for all demand points (Daskin, 2008). This model is often used when there are not enough facilities while the facility has to serve all the patients in a specific area (Peng and Afshari, 2014). The p-centre model is suitable when equity for every patient is essential (Du and Zhou, 2018). This model is often used in emergency health care planning (Hochbaum and Pathria, 1998). The set covering problem can be used in a p-centre problem when the location of the facilities is restricted to nodes of a network (Marianov and Taborga, 2001).

P-centre location problems are the third classical type of covering-based problems, which minimise the maximum travel distance (or time) among all demand points and the allocated facilities, considering that every demand point is covered. When the facilities are uncapacitated, the demand points are assigned to the nearest open facilities.

Median based models aim to minimise the demand-weighted average distance or distance cost between a demand point and the assigned facility. These locations are referred to as the medians of the network. These problems are also referred to as location-allocation problems as they try to solve facility location and demand allocation simultaneously. These models are often used in distribution planning, where minimising the transportation cost is essential. This class of discrete location models includes the p-median and fixed-charge model (Daskin, 2008).

The p-median model aims to minimise the total travelling distance from the demand node to the nearest facility by locating a number (p or less) of facilities. With an increase in average travelling distance, the facility's accessibility decreases and, therefore, the location's efficiency decreases. An assumption made by this model is that all patients choose the nearest facility. This model is one of the most popular, especially in locating public facilities (Rahman and Smith, 2000). Some drawbacks of this type of model is that it may be inequitably forcing some users to travel very far, and it does not consider the capacity of the facility (Meskarian et al., 2017). When working with a median distance, having a few households travelling very far to reach a facility is cancelled out by most of the households that have short distances to travel since the median is used and not the mean.

Fixed-charge facility location problems are similar to p-median problems. The main difference is that fixed-charge facility location problems attempt to minimise the total travelling cost and the cost of opening the facilities. In contrast, p-median problems disregard the cost difference at different candidate locations. All problems that do not fit the abovementioned categories belong to the last category (other).

With an understanding of the various facility location models, the focus was placed on facility location models in the primary health care environment.

2.2.1 Primary health care facility location problems

Primary health care is the first contact care given in hospitals and clinics. In this project, container clinics, which are part of primary health care, are investigated. An investigation was done into primary care facility location problems.

Selecting the criteria or objectives is a crucial aspect of location models. The objectives for private and public facilities differ. Private entities make location decisions intending to increase the profit margin (Rahman and Smith, 2000). In contrast, for primary (public) health care facilities, ease of access for the community is one of the critical considerations (Ahmadi-Javid et al., 2017). In recent years, cost minimisation has become a popular objective in primary health care facility location problems.

When looking at the types of problems investigated in the primary health care environment, more than half of the literature is on median-based location problems. Covering-based location problems have contributed to a large portion of the literature in the last decade (Ahmadi-Javid et al., 2017).

Mestre et al. (2015) applied a p-median model in which the distance to a health care facility was set as the constraint in an uncertain demand environment. Beheshtifar and Alimohammadi (2014) defined their p-median problem with 4466 demand points and 100 candidate sites to find the optimal locations for clinics while minimising the transportation cost and land cost. Kim and Kim (2013) determined the location of public health care facilities within a given budget and by maximising the number of patients served in both private and public facilities, using a Lagrangian heuristic for their p-median problem. Das et al. (2020) also made use of p-facility location model to place new facilities amongst existing facilities while minimising the total transportation cost. citetDzator2019 did a comparison of optimisation models for placing ambulance stations in Queensland, Australia and found that the p-medial model provides better solutions than the maximal covering models since it is not dependant on the predetermined weights of the maximal covering models.

The problem solved in this report is similar to these p-median problems, intending to minimise cost and improve accessibility. In the majority of these p-median health care-related problems, either transportation cost or building costs are minimised; for this model, building costs, transportation costs and distribution costs are minimised.

Another critical distinction in facility location models is capacitated versus uncapacitated. With capacitated facility location models, each facility has a specific capacity or limit, which it cannot exceed without penalties. When using uncapacitated facility location models, there is no limit set on the capacity of the facility; therefore, all facilities can serve an infinite demand (Sun, 2012). Many location-allocation problems in the literature incorporates a capacity constraint. Beheshtifar and Alimohammadi (2014); Güneş et al. (2014); Graber-Naidich et al. (2015); Shishebori and Yousefi Babadi (2015) defined capacitated problems, limiting the maximum service capacity. Irawan et al. (2020) also incorporated capacities in their p-median problem by optimally locating facilities with optimal capacities as well. By including the capacity in the problem, these papers investigate more than the accessibility; they also investigate the availability.

The problem addressed in this report has similarities with many of the problems solved in literature defined as uncapacitated p-median problems. The objective is to minimise the total cost, with clinics being within a 5 km radius for all residents of the areas investigated. The clinics in this report have a maximum capacity that they can serve. Therefore, the problem in this report can be defined as an uncapacitated p-median problem with uncertainty.

2.2.2 Facility location decisions under uncertainty

Strategic planning of a health care network involves long-term decisions like facility location. These decisions need to be robust to face future demand and supply pattern changes. The uncertainty in the environment increases as the planning horizon moves forward (Mestre et al., 2015).

Rosenhead et al. (1972) divided decision-making environments into three categories: certainty, risk, and uncertainty. In certainty environments, all the parameters are deterministic and known; this is not the case for the risk and uncertainty environments. In risk environments, the uncertain parameter values are governed by known probability distributions. Problems in the risk environment are usually solved using stochastic optimisation. Problems in an uncertain environment are solved with robust optimisation. Problems are considered uncertain when the probability distributions of the random parameters are not known to the modeller. Both stochastic and robust optimisation aims to find a solution that will perform well, given any possible realisation of the random parameters. The definition of well differs from modeller to modeller and application to application. Defining the appropriate performance level and measures is part of the modelling process (Snyder, 2006).

In the case where the probability distributions of the uncertain parameters are unknown, the parameters are restricted. They lie in a pre-specified interval for continuous parameters or scenarios are developed using discrete parameters. When dealing with scenarios, there are two common drawbacks. Developing these scenarios can be a daunting task since nobody knows what will happen in the future and there is an unlimited number of possible scenarios. Detailed planning documents and expert opinions are often required to develop plausible scenarios to use (Kchaou Boujelben and Boulaksil, 2018; Snyder, 2006). The task of developing scenarios is less daunting when using municipal planning documents. The second drawback is that a limited number of scenarios can be considered due to computational restrictions. However, the scenario approach allows for more controllable models and allows parameters to be statistically dependent, which is not always possible with continuous parameters described by probability distributions (Snyder, 2006). This statistical dependence is required in the model since the demand is correlated across geographical regions and time periods.

Two common approaches in scenario modelling are the min-max cost approach and the min-max regret approach. The min-max cost approach aims to minimise the maximum cost across all the scenarios. This is a very conservative approach emphasising the worst-case scenario, therefore producing inadequate solutions for the other scenarios. This model is best suited for problems when the solution should function well, even in the worst-case scenario. Often it is more practical to plan based on a fractile target than the worst-case scenario. For example, when a hospital tries to meet all the demand 90% of the time and therefore risks turning away patients in extreme cases such as pandemics (Snyder, 2006).

A less conservative alternative approach was developed by Daskin et al. (1997): the α-reliability method. This approach ensures that the set of scenarios used have a probability of occurring at least α. By including the probability of the scenarios in the model, a more realistic solution can be provided (Mestre et al., 2015). In the regret approach, the difference between the cost of each scenario's optimal and the cost of the solution in a given scenario is minimised.

Robust optimisation has been applied in various areas such as contracts in supply chain (Gumte et al., 2021), electricity generation (Yang et al., 2021) and distribution networks (Xu et al., 2021). Baron et al. (2011) used robust optimisation to locate multiple facilities in a network with uncertain demand over multiple periods. The locations and capacities of the facilities are determined and the allocation of demand to these facilities is determined in the study. The study consisted of 15 nodes and 20 periods. Baron et al. (2011) concluded that a problem of this size would be difficult to solve using other stochastic methods. In robust optimisation, the linear programme is converted into its robust counterpart by replacing the constraints with uncertainty coefficients to reflect the uncertainty.

When dealing with multiple objectives, which often occurs in robust optimisation, goal programming can be used to find the most satisfactory solution in a feasible region (Chang, 2015). This technique has goals (desired values) for each objective; the model then finds the solution that gets the modeller as close as possible to these desired values. Goal programming minimises the deviation between the goal for each objective and the value of the objective function.

Goal programming was used by Wichapa and Khokhajaikiat (2017) to find the best location for infectious waste disposal from three candidate facilities serving 40 hospitals in sub-Northeastern Thailand. The total cost (facility cost, operating cost, and transportation cost) of the solution was minimised and the priorities' weights were maximised. Miç et al. (2019) used a mixed integer weighted goal programming model to locate 42 primary health care clinics out of 77 candidate sites in the north of Idleb. For the study, the total cost was minimised and the number of people that have access to a clinic was maximised with a limited budget. The study focused on facility location in conflict areas, but the framework can be used in different situations.

Ghodratnama et al. (2015) used goal programming in an uncertain environment to locate distribution hubs for a two-echelon network. The uncertain parameters in the model were transportation costs, demand and opening and closing costs. The objectives that were minimised are the total transportation and installation cost, the service times, and the greenhouse gases. Goal programming assisted the authors in successfully finding robust locations for the distribution hubs given the uncertain environment. The problem investigated in this report also seeks a robust solution in an uncertain environment; goal programming is an efficient solution method for robust optimisation since the problem has various conflicting objectives. The first objective of minimising the total building cost will minimise the number of clinics placed. The second objective of minimising the travel distance to the clinics for the households will push to as many as possible clinics as close as possible to the households for all three scenarios. However, the households are not located in the same locations for all the scenarios. The last objective of minimising the distribution distance will push the model to locate as few as possible clinics as close as possible to the distribution centre.

There is a lack of dynamic location models in the literature that consider the changes in the problem environment over time, such as patient population and population migration (Ahmadi-Javid et al., 2017). This project aims to assist in closing this gap in the literature, using robust optimisation and goal programming to locate health care facilities in an uncertain environment using multiple scenarios.

2.3 Solution generation methods

A variety of tools and techniques are available to solve facility location problems. Exact methods or approximate methods can be used. If the network is relatively small, exact methods can be used to find optimal solutions. These methods are sure to find an optimal solution to the problem. If the problem gets too large, exact methods are no longer feasible and approximate methods have to be used to obtain a reasonably good solution in a reasonable time (Talbi, 2009).

Facility location problems are considered NP-hard problems. The NP-hardness of the problem is increased with the size of the network. The aggregation level influence the size and the NP-hardness of the problem (Cebecauer and Buzna, 2017). The network may have thousands of nodes, where each node represents a block or a zone in a city. Since robust multi-facility location problems are classified as NP-hard problems, heuristics was the focus of the research for the solution generation methods (Clarke and Wright, 1964).

Heuristics are used to find good solutions or a satisfactory feasible solution to the problem. A common heuristic used in facility location problems is local search. This method proposed by Kuehn and Hamburger (1963) starts with a feasible solution set and iteratively improves the solution by moving to the best neighbouring solution. The neighbours of a feasible solution in facility location models are often obtained by adding or removing a facility or changing the facility's location. Kim and Kim (2013) used a heuristic algorithm based on the Lagrangian relaxation to locate public health care facilities with a constrained budget while maximising the number of patients served. This study was done for a small case study of 33 demand points; therefore, a good solution was reached in less than 13 seconds. The problem addressed in this report is much larger than the problems solved with heuristics and, therefore, metaheuristics are investigated.

Metaheuristics serve two primary purposes: solving problems faster and solving large problems. Various metaheuristics are available. In recent years, Genetic Algorithm (GA)s, Tabu Search (TS), Simulated Annealing (SA) have been most applied in facility location problems and are therefore further investigated in this section (Beheshtifar and Alimohammadi, 2014). After an overview of these metaheuristics and an investigation into their application in the facility location environment, a critical review is done to select the most appropriate metaheuristics to use for this project.

2.3.1 Genetic algorithm

A GA is an evolutionary algorithm based on the adaptive process in nature. GAs take two possible solutions and change aspects of the solutions to develop two new possible solutions. This process is done randomly to diversify the solutions space. As the algorithm evolves, the better solutions have a higher chance of being selected again (Katoch et al., 2020). The pseudocode of a typical GA as presented by Reeves (2003) can be seen in Algorithm 1. A GA works with a set of solutions (population), with elements represented by individuals (chromosomes). The first step is to select an initial population of chromosomes. These chromosomes evolve by applying crossover and mutation operators. A new population is created each time these operators are applied. The new population is evaluated against a fitness function to determine how good the solution is for the problem. In most cases, this function is the objective function of the problem.

Typically, the larger the population size, the better the final solution. However, with an increase in the population size, the computation time increases proportionally (Rajagopalan et al., 2007). Shariff et al. (2012) used a GA to locate health care facilities using the maximal cover model. Aytug and Saydam (2002) applied a GA on a maximum expected coverage location problem to locate ambulances. GAs produce high-quality solutions for set cover location problems citepgazani2021capacitated. Beheshtifar and Alimohammadi (2014) used a GA to find good sites for new clinics while minimising the total travel cost, inequality in access to the clinics and the total cost of building the clinics. The length of the chromosome in this project varied depending on the number of clinics to be opened. Different types and levels of relative importance for each objective were considered to obtain the best solution for the case investigated in Tehran, Iran. This is a similar problem to the problem addressed in this report. However, in this report, the placement is considered over 12 years and not only in a static environment. For the problem investigated in this report, the individual is a list of all the possible sites. Each instance will represent a zone and whether a clinic is opened in the zone in a given year.

Algorithm 1: Genetic algorithm

```
Create initial population of chromosomes;
while stopping condition is not satisfied
    repeat;
    if Crossover condition is satisfied
        Select parent chromosomes;
        Define parameters to crossover;
        Apply the crossover operator;
    end
    if Mutation condition is satisfied
        Choose mutation points;
        Apply the mutation operator;
    end
    Evaluate fitness of offspring;
    Until Sufficient offspring is created
end
return The best individual found
```

2.3.2 Tabu search

A TS is based on the local search heuristic that restricts the feasible neighbourhood by neighbours that are excluded. This metaheuristic has a unique feature called a memory list. This list ensures that once a solution is entered into the list, the solution cannot be revisited for some time (Rajagopalan et al., 2007). This short-term memory, tabu list, stores recently visited solutions (or attributes of recently visited solutions) to avoid being stuck in a local optimum. If all the neighbours are tabu, a move that worsens the objective value is accepted. The search stops after a fixed number of iterations or after several consecutive iterations have been performed without any improvement to the best-known solution (Gendreau, 2008). The pseudocode for a TS is shown in Algorithm 2, as described by Talbi (2009). TS is a common heuristic used in facility location modelling.

Since the TS only uses one vector to search the space, the quality of the solution is highly dependent on the quality of the initial vector. The best solution of the TS is not necessarily the final solution of the last vector since the solutions are stored in the memory list (Rajagopalan et al., 2007). This memory list stores previous vectors and their solutions to avoid cycling and keep the best vectors and their solutions. Once the algorithm has reached its stop criteria, the memory lists can be viewed to find a good solution if the final solution was not good. The TS algorithm has been successfully applied in numerous domains, including health care. Klein et al. (2020) used TS to near optimally identify a network of dialysis facilities in rural areas. They were able to determine a good network of facilites with a budget and capacity constraint. Sun (2012) used a TS algorithm to solve numerous capacitated facility location problems. For small problems, the TS was able to find the optimal solutions. For larger problems with more than 1000 nodes, the solutions were not the optimal solutions, however, they were close to optimal. Sun (2012) concluded that TS performs well for capacitated p-median problems. Since the problem addressed in this report is a p-median problem, the TS is a strong contender for the metaheuristic to be used.

Algorithm 2: Tabu Search

s = s_0 ;
Initial solution Initialise the tabu list, medium-term and long-term memories;
Repeat
Find best admissible neighbour $s\prime$ non-tabu or aspiration criterion holds ;
s = $s\prime$;
Update tabu list, aspiration conditions, medium- and long-term memories ;
if *intensification criterion holds*
    intensification;
end
if *diversification criterion holds*
    diversification;
end
Until Stopping criteria satisfied
return Best solution found

2.3.3 Simulated annealing

The SA optimisation process is based on the heating and cooling of metals known as the annealing process. Similar to the physical process, the results of the algorithm are gradually improved until a good solution is identified. A SA algorithm can accept worse solutions, with some probability, as a mechanism to escape a local optimum. This is a very time-efficient algorithm to run to get a reasonably good solution (Albright, 2007). The pseudocode for a SA is shown in Algorithm 3, as described by Talbi (2009).

As in the TS, the SA algorithm uses a single vector to search for a solution. With this algorithm, the quality of the final solution is also dependent on the initial vector. Rajagopalan et al. (2007) found that the SA algorithm gets stuck in local optimas more frequently than any of the other metaheuristics when working with large data sets of 1024 zones or more. The study found that setting the right temperature is crucial for not getting stuck in a local optimum. Levanova and Gnusarev (2018) used a SA algorithm to solve a p-median problem with elastic demand

and a cost constraint. The algorithm was successfully used to place supermarkets and hypermarkets to attract the biggest demand share. Syam and Côté (2010) used SA to determine good location for traumatic brain injury units. In the study, 100 candidate locations were considered, and 15 treatment units were opened while minimising the total cost and service proportion requirements. The computation time increases exponentially as the number of candidate locations increases. This is a crucial correlation to note as it is a large problem under investigation in this project. Chiyoshi and Galvao (2000) tested a SA algorithm on multiple uncapacitated p-median problems with a total number of vertices ranging from 100 to 900. The study found that the total number of vertices and the number of facilities to be located have a strong correlation with the run time. An increase in the number of vertices or the number of facilities to be located leads to an exponential increase in the run time.

Algorithm 3: Simulated annealing

Input: Cooling schedule
s = s_0 (Generation of the initial solution)
T = T_{max} (Starting temperature)
Repeat
Repeat (At a fixed temperature)
Generate a random neighbour $s\prime$;
$\triangle$E = f($s\prime$) - f(s) ;
if $\triangle E \leq 0$
    s =$s\prime$;
else
    Accept s with a probability $e^{\frac{-\triangle E}{T}}$
end
Until Equilibrium condition
T = g(T);
Until Stopping criteria satisfied (e.g. T < T_{min})
return Best solution found

A comparison of these three algorithms was done by Arostegui et al. (2006), focusing on facility location problems. These algorithms were evaluated on two aspects, the time to reach a good solution and the quality of the solution. Regarding the time to reach a satisfactory solution, it was found that TS is the fastest, followed by SA and then GA.

Rajagopalan et al. (2007) also did a comparison of these three algorithms and found that TS and SA have faster times to a solution. However, all the investigated metaheuristics solved the larger problems within minutes. Therefore, the time difference can be considered insignificant. When looking at the quality of the solutions between the methods investigated in the study by Rajagopalan et al. (2007), the quality of the GA solutions were better than the quality of the other two algorithms compared to the exact method solutions. This could be due to the fact that the quality of the GA solution is not as dependent on the quality of the initial solution as with the TS and GA. Since the time difference of the algorithms is considered insignificant and the quality of the GA is better than the rest and not dependent on the initial solution, the GA was selected as the heuristic to solve the problems in this report.

2.4 Concluding remarks

In the literature, many household attributes were found to correlate with health care demand and utilisation. The main attributes are age, gender, income and distance to the health care facility. The research done on facility location models led to the definition of the problem. Based on the literature, the problem of this report can be classified as a p-median problem in an uncertain environment. Methods to deal with uncertainty in facility location models were investigated. Robust optimisation was identified as the best solution method to deal with the uncertainty faced in this problem. Scenarios and goal programming were selected as tools to assist in the robust optimisation. Finally, possible solution methods were investigated. As this is an NP-hard problem, several heuristics and metaheuristics were considered. Based on comparisons found in the literature, the GA was selected to solve the problem. Based on the information gained from the literature review, a detailed methodology is developed in Chapter 3.

Chapter 3

Methodology

This chapter unpacks the methodology employed to determine robust locations for the container clinics, given three possible scenarios for the three metros. The methodology includes the unpacking and analysis of the data, conversion of household attributes into health care demand and model development.

A brief outline of the methodology is explained in Figure 3.1. The first step was data preparation, followed by solution generation. The last step was to conduct additional analyses.

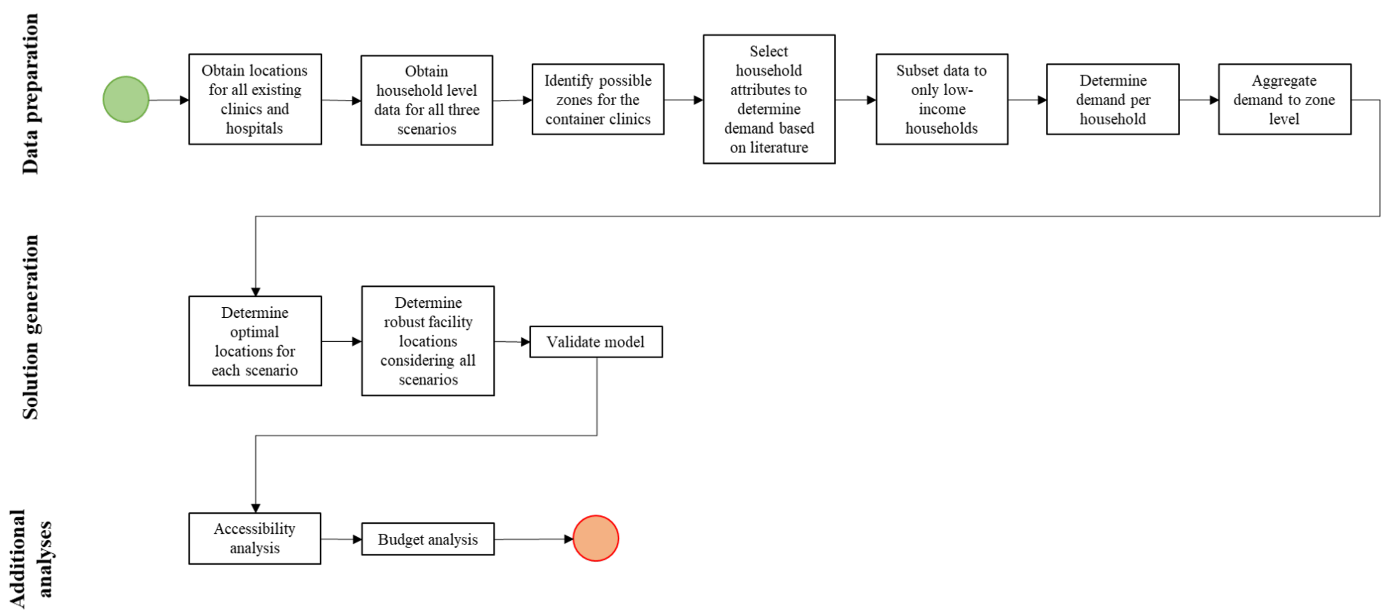


Figure 3.1: Road map to identify robust locations for container clinics given multiple scenarios

3.1 Data unpacking

For this project, two data sets are required. The first is a data set of all the existing public clinics and hospitals in the three metros under investigation. This data is required to calculate the initial accessibility measures. The existing public health care facilities are used as the base facilities and all the container clinics built are added to this set. The second data set required is the synthetic household distribution and attribute data. The household data is necessary for the primary health care demand calculation in the model. The process of obtaining these data sets is described in this section.

3.1.1 Public health care location data

The distribution of existing hospitals and clinics as per the Department of Health (2018) are illustrated in Figure 3.2. Johannesburg has 238 existing clinics, Ekurhuleni 174, and Tshwane 146. Even though Tshwane is the largest metro, it has the least number of clinics. In Tshwane, 42% of the population is further than 5 km from an existing clinic. In Ekurhuleni and Johannesburg, these percentages are much lower, 17% and 12% respectively. Based on this, there is a clear need for primary health care facilities closer to the residents in all three metros.

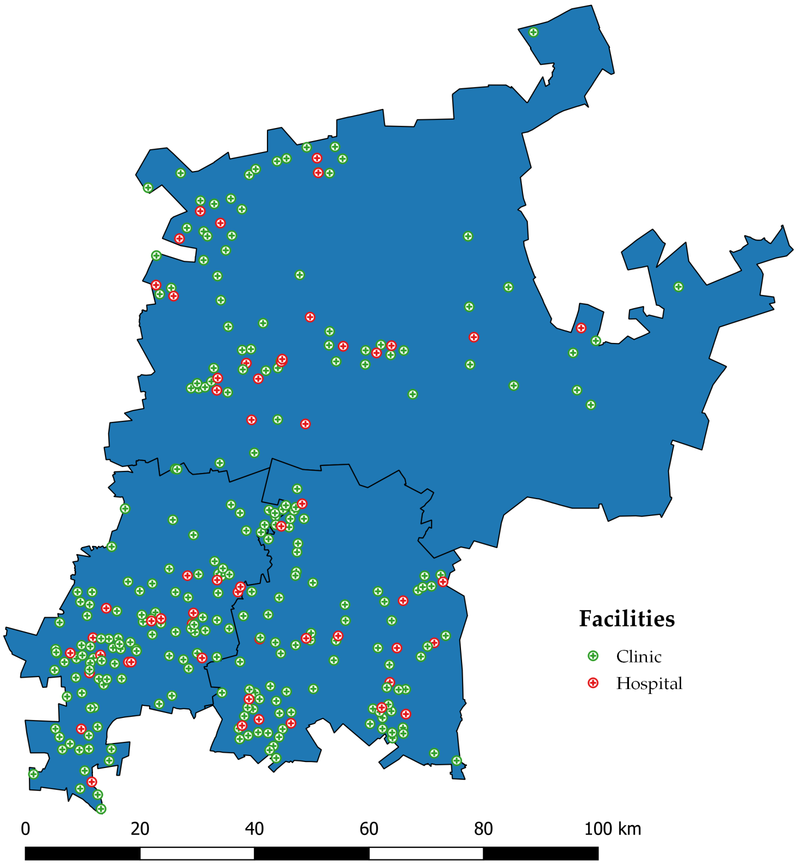


Figure 3.2: Distribution of hospitals and clinics across the three metros

3.1.2 Scenario and household data

To identify vacant land, the combined area of all three municipalities were divided into 28 461 mostly homogeneous square zones of approximately $1km^2$ in size, as illustrated in Figure 3.3. Each zone comprises several parcels, with varying size. Parcels refer to cadastral parcels — a representation of the individual erven. Parcels are classified according to their underlying land-use and a parcel could either be built-up (having one or more buildings present) or vacant (having no buildings). The built-up parcels can further be classified as commercial or residential, depending on the building use. The difference between vacant and built-up parcels can be seen in Figure 3.3. The dark blue dots represent commercial buildings and the light blue dots residential buildings. The parcels with the dots are built-up and cannot be developed. The vacant parcels can be developed and, therefore, only these parcels are considered when calculating the vacant area in the zone. Using this received zonal and parcel data, the candidate zones were identified. For this project, only vacant zones that are 35 m^2 or larger were considered as candidate locations for the container clinics.

Urban growth modelling is an interdisciplinary field that encapsulates both scientific and technical research areas. Various interrelated aspects are investigated simultaneously, such as transport models, geographical information science, urban geography and complexity theory. The essence of urban modelling can be described as a process of encoding part of the real world into a model system, running the model a number of times to simulate

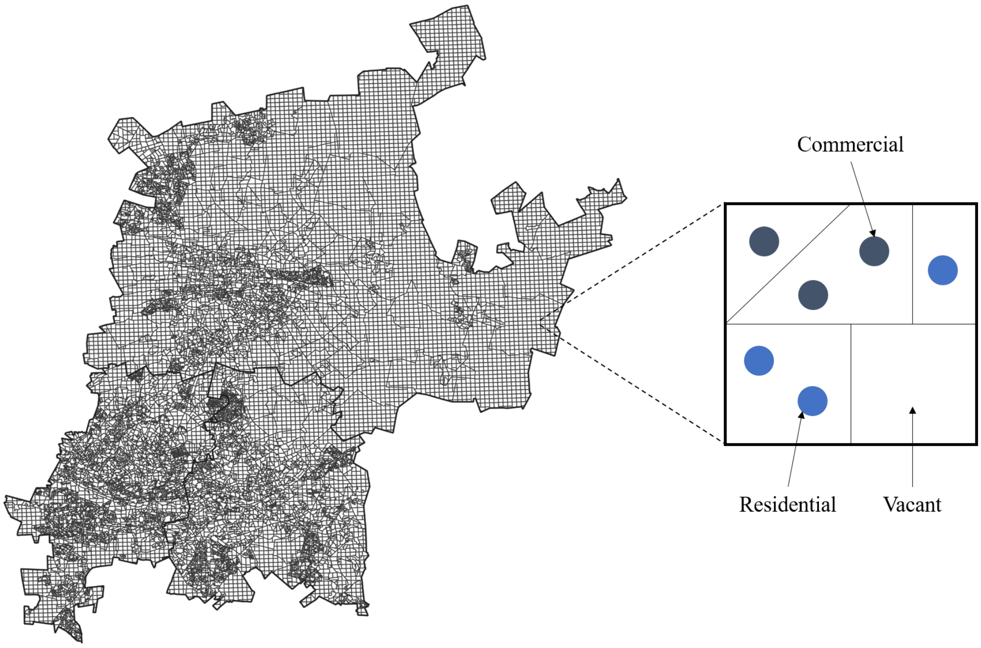


Figure 3.3: The three metros divided into zones and a zone divided into parcels to represent vacant and built-up parcels

specific policy scenarios 30 years into the future, and then evaluating the relative success of each policy scenario in achieving the stated objectives or performance measures of a city or region. The urban growth simulation model developed and used by the Council for Scientific and Industrial Research (CSIR) aims to support cities in land use planning and the optimisation of the location of facilities and services.

Three possible city forms and population distributions for the three metros were developed by the CSIR using an urban growth simulation model. The scenarios reflect differences in household and economic growth in the metros. The three scenarios are called: *Trend*, *Economic spike* and *Relocation*. These scenarios were developed using various municipal planning documents such as the Metropolitan Spatial Development Framework (MSDF), Integrated Development Plan (IDP), Built Environment Performance Plan (BEPP) and the Integrated Transport Plan (ITP). These documents were used to identify development projects in the metros and the importance of the projects to the municipalities.

All these scenarios started with the same synthetic population distribution in 2018. The distribution of households in 2018 can be seen in Figure 3.4. In 2018, there were 2 858 933 households in the lower-income categories being investigated in this project. A description and unpacking of each scenario follow.

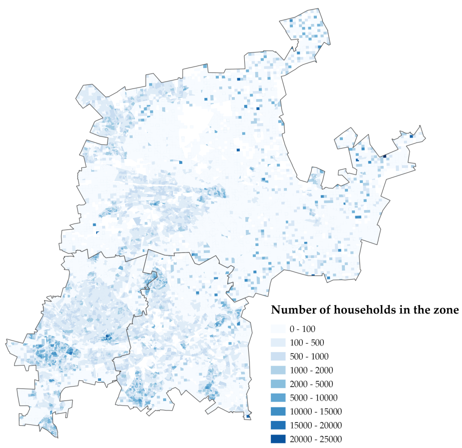


Figure 3.4: The distribution of households across the metros in 2018

Trend scenario

The trend scenario is a continuation of the current development trend. The household growth and economic development follow the current growth trends based on historical data from 2011 to 2018. Half of the planned municipal development projects will be implemented in the years up until 2030. Figure 3.5 shows the distribution of lower-income and lower-middle-income households by 2030. In this scenario, the households that fall into the lower-income categories increased to 5 932 592 by 2030.

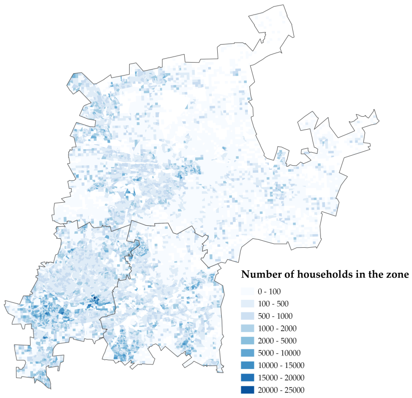


Figure 3.5: The distribution of households in the trend scenario in 2030

Economic spike scenario

The economic spike scenario is based on the trend scenario. The development follows the current trend, with one exception, the development focus is on employment opportunities. This scenario sees an increase in the number of jobs available and a decrease in the number of low-income and lower-middle-income households as a result of an increase in jobs. The distribution of these households is displayed in Figure 3.6. There as significantly fewer households in the investigated income classes in this scenario when compared to the trend scenario. In this scenario, by 2030, the number of households earning R108 000 or less per year is 3 589 217.

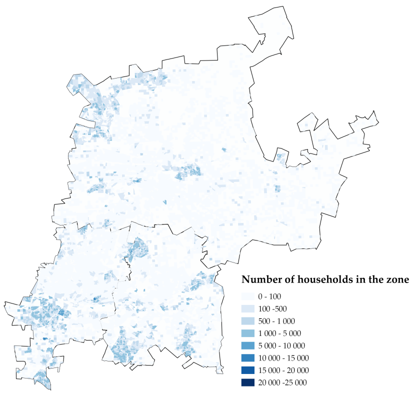


Figure 3.6: The distribution of households in the economic spike scenario in 2030

Relocation scenario

This last scenario is once again based on the trend scenario. However, more housing development projects are implemented, allowing households to relocate to these three metros from rural areas. In this scenario, the number of households in the lower-income and lower-middle-income category increased. There are 7 563 688 households in the lower-income and lower-middle-income categories by 2030. In Figure 3.7, the distribution of these households is illustrated. The increase in households is mainly in the Tshwane municipality.

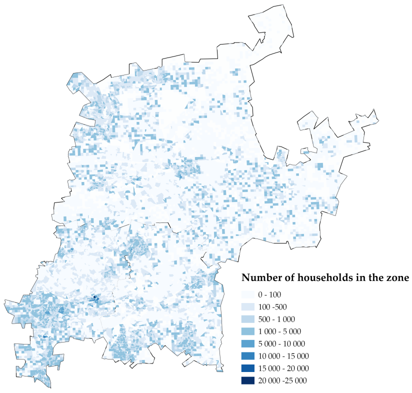


Figure 3.7: The distribution of households in the relocation scenario in 2030

For each of these scenarios, household attribute tables are available up until 2030. These attribute tables are unpacked in the data preparation and used as input for the individual scenarios. Further analysis of the scenarios with regard to the specific cases follows in the next section.

3.2 Data preparation and analysis

3.2.1 Identify vacant zones

With the public health care data and synthetic household information available, data preparation commenced. For each of the scenarios, the zones with enough vacant space for a container clinic were identified. The dimensions of the containers used are 12 m in length, 2.3 m in width and 2.4 m in height. Based on the typical container footprint covering approximately 28 m^2, vacant land of at least 35 m^2 is deemed sufficient to locate a typical container clinic (Cooke et al., 2010). To determine the vacant area in a zone, the area of all vacant parcels in the zone were aggregated.

All zones with enough vacant space for container clinics were included in the set of candidate locations. The location of these zones within the three metros are indicated in blue in Figure 3.8. A pharmaceutical distribution centre was identified in Centurion and was used in this study. The distribution centre is illustrated with the red dot in Figure 3.8. All deliveries to the clinics are made from this distribution centre. This distribution centre was selected because it is an existing pharmaceutical distribution centre close to the centre of the three municipalities and many of the other existing pharmaceutical distribution centres are located in other provinces. By selecting a distribution centre in another province, the total distribution distance would increase significantly. The long distance travelled to reach Gauteng from another province would make the shorter distances travelled within the province to the clinics seem insignificant.

3.2.2 Demand conversion

The next step was to identify the available household attributes relevant to health care demand. For each scenario, household attribute tables are available up until 2030. The household attributes available for the UrbanSim model are listed in Table 3.1. Based on investigated literature, the household income, the number of children, distance to nearest health care facility and household size can be used to determine health care demand.

The following three attributes were used with proportions: the number of children, income, and distance to the nearest facility to convert the data into health care demand. For all households, the probability of visiting the clinic if it is within a 5 km radius is one. This probability decreases exponentially with the increase in distance. Following the distance decay functions used in the literature, the distance decay function used for this project was calculated as $y = 0.95^x$, where x is the distance (in km) to the nearest clinic and y is the probability of visiting a clinic when ill. The scale used for this conversion is shown in Figure 3.9. In the model, the distance to the nearest facility is a dynamic variable as the distance to the nearest facility depends on where the facilities are placed. Figure 3.10 shows the proportion of the lower-middle-income households in each distance bracket for the

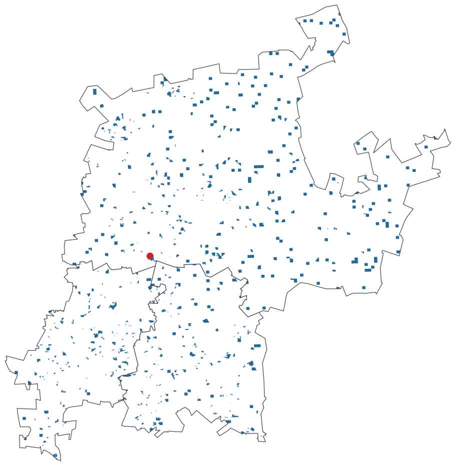


Figure 3.8: Locations of candidate zones in the three metros and the location of the distribution centre

Table 3.1: Household attributes available

Explanatory variable	**Variable type**
Number of children in the household	Numerical
Number of workers in the household	Numerical
Household income per year	Numerical
Age of household head	Numerical
Household size	Numerical
Distance to nearest health care facility	Numerical

base year (2018) and the three scenarios, where these distances are the distance to the nearest existing clinic. The proportions in the trend scenario are slightly higher than in the base year. In the economic spike scenario, the proportions are slightly lower than the base year. Lastly, the relocation scenario proportions are slightly higher than those of the trend scenario. All these variations in proportions are simply due to the change in the number of households in the investigated income categories.

Only lower-income and lower-middle-income households were investigated in this study. A total household income of R 108 000 per year was the threshold used to differentiate lower-income and lower-middle-income households from others. These lower-income and lower-middle-income classes were subdivided into income classes used in the simulation model. For the rest of the report, the following classes are used: class 1 (R 0 - R 9 600), class 2 (R 9 601 - R 42 000) and class 3 (R 42 001 - R 108 000). Each of the income classes had a probability of an individual visiting a clinic when ill assigned to them. In line with findings in the literature, these probabilities decrease as the income decreases. Exact proportions for a South African context were not found during the literature study, therefore the proportions used in this study were selected using the knowledge gained from the literature and applying it in the South African context. The probabilities used in this study are shown in Table 3.2.

Table 3.2: Probability of individual visiting a clinic when ill based on annual household income

Income class	**Household income per year (R)**	**Probability of visiting the facility when ill**
1	0 - 9 600	0.6
2	9 601 - 42 000	0.7
3	42 001 - 108 000	0.75

A comparison of the number of households in each of these categories for all the scenarios is illustrated in Figure 3.11. The distribution of households in the income categories remained more or less the same in all the considered scenarios; only the number of households in

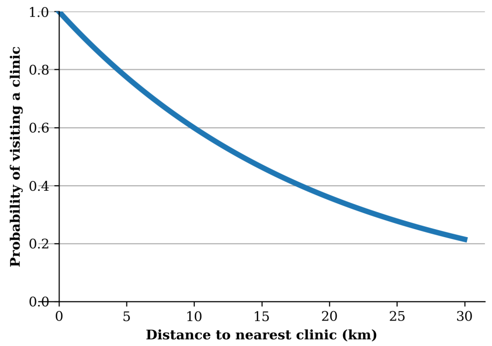


Figure 3.9: Distance decay function used

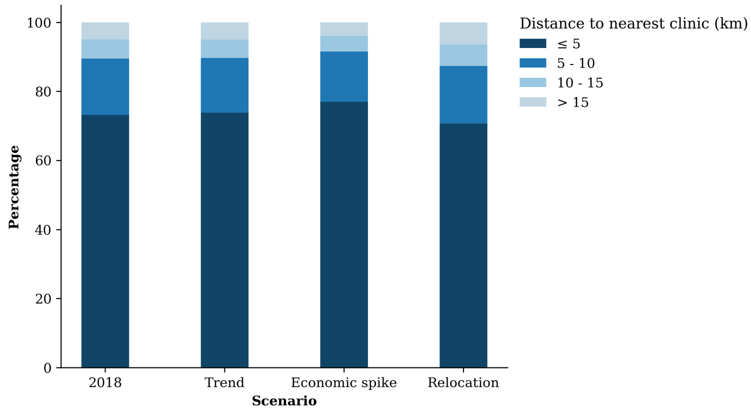


Figure 3.10: Accessibility breakdown per income category for the three scenarios

the income classes investigated in the economic spike scenario was significantly lower than in the trend scenario. However, it was almost the same as in the base year. In the relocation scenario, there is an apparent increase in the number of households in the lower-income classes compared to the trend and the economic spike scenario.

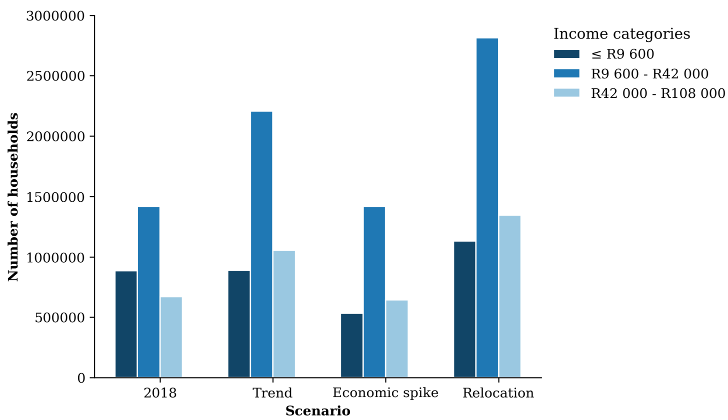
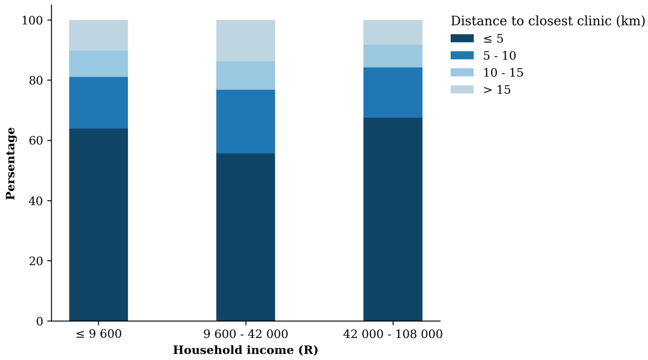


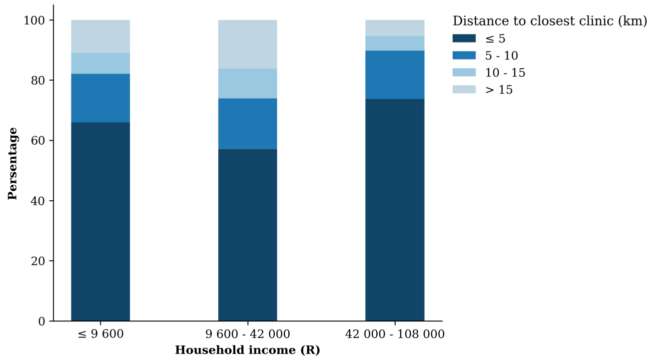
Figure 3.11: Income distribution comparison of the base year and three scenarios

Almost 40% of the population in these metros form part of the study based on the yearly household income of R 108 000 or less. A breakdown of the income and distance to the nearest clinic for the base year and all three scenarios only considering existing clinics is illustrated in the graphs in Figure 3.12. In the base year, the trend scenario and the relocation scenario, income class 2 has the highest percentage of households further than 5 km away from a clinic. For the economic spike scenario, it is income class 1 that has the highest percentage of households further than 5 km from the nearest clinic. This percentage is significantly higher than in any of the other scenarios. The trend scenario graph looks similar to the base year graph, however, there is a reduction in the percentage of households not within a 5 km radius of a clinic. About 40% of the households in the bottom two income classes of the trend and relocation scenario are further than 5 km from the nearest clinic. In the relocation scenario, access to clinics is spread more evenly across the income classes than in the other scenarios. These graphs indicate the need for health care facilities closer to the community.

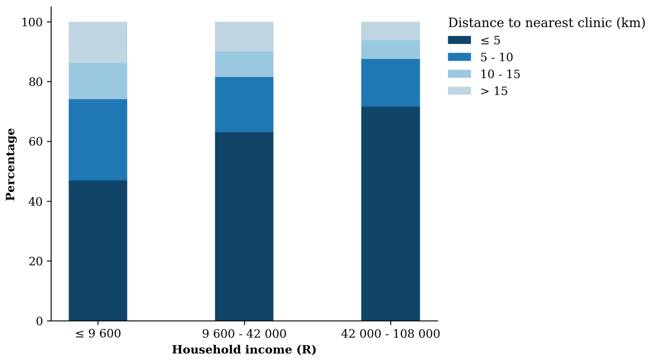
From the literature, it was identified that households with children are more likely to visit a health care facility. Therefore, in this study, households with children are considered more likely to visit a health care facility than households with no children. If a household has one or more children, the probability of visiting a health care facility when ill is 0.85. For a family with no children, this probability reduces to 0.75. Exact probabilities for a South African case study were not found, therefore these arbitrary values were selected based on the research done during the literature study.



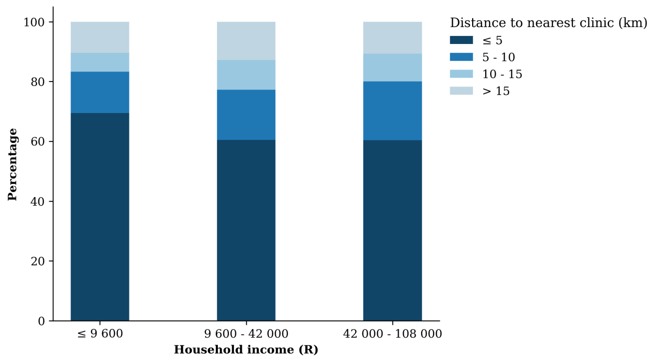
(a) Base year



(b) Trend scenario



(c) Economic spike scenario



(d) Relocation scenario

Figure 3.12: Distance breakdown to existing clinics per income class for the base year and the three scenarios

In order to determine the health care demand per household, these probabilities were multiplied by the household size. This demand for each household was aggregated to a total demand per zone used in the model. Algorithm 4 depicts the pseudo algorithm used for demand conversion.

Algorithm 4: Health care demand per zone calculation without considering distances

for *year in range(start year, end year)* **do**
    Calculate probability of visiting a clinic based on household income (p_1)
    Calculate probability of visiting a clinic based on the number of children in the household (p_2)
    Calculate demand per household: $(household\ size * p_1) * (p_2)$
    Add the demand per household for all the households in the zone
end
return Demand per zone without considering distance

3.3 Model development

Once all the data sets have been prepared, the model is developed. The individual scenario models, as well as the robust model, are developed. The model development is fully described in Chapter 4.

3.4 Verification and validation

Once a good configuration for each scenario and the robust configuration have been determined, the model verification and validation is done. The verification of the model is simply to confirm that the model does what it is supposed to do: find good scenario configurations and robust configurations. The validation determines if the solutions from the model are valid. For this project, the robust configuration is deemed valid if the total difference between the robust configuration and scenario configuration is no more than 25% for each of the scenarios. No fixed percentages were found to determine if a solution is robust or not as it is subjective to every case, therefore 25% was selected since the three scenarios being investigated are very different. This robustness level can be adjusted based on the similarity required in the solutions. The smaller the robustness level, the closer the solution has to be to the original solutions.

With the methodology set out for this project, the last step was to select the tools to solve the problem.

3.5 Tools

A variety of software packages are available to solve the problem stated above. Standard software packages used to solve these facility location problems include Lingo, R and Python. Lingo is a tool used to solve linear, non-linear, and numerous other optimisation models. Python is a common programming language and it is often used for optimisation models. R is also a programming language that can be used for solving the problems identified in this report. Lingo has a free version with limited capabilities, however, with the scope of the problem, the full version will be required. Both Python and R are open-source software. Any of these software packages could be used, however, for compatibility reasons with the UrbanSim models, Python was selected as the tool to use for the analysis models.

There are numerous packages available in Python to assist in solving the problem. A few of these optimisation packages were investigated. Pyomo (Hart et al., 2017), DEAP (Fortin et al., 2012) and PuLP (Mitchell et al., 2011) are the alternatives investigated for this project. PuLP is an open-source package used for linear programming. Pyomo is an open-source Python package for formulating and analysing optimisation models. This package is not restricted to linear programming; it can handle a wide range of problems, including quadratic and mixed integer stochastic programming. DEAP is an evolutionary computational framework package that can be used to solve a wide range of problems using evolutionary algorithms. Due to the size of the problem, heuristics such as evolutionary algorithms will be required to solve the model in an acceptable time. DEAP and Pyeasyga are both heuristic packages that can be used in Python, however, DEAP has more alternatives within the package and more documentation for ease of use. Therefore, DEAP was the selected Python package to implement to solve the problem.

3.6 Concluding remarks

The methodology followed in this project was laid out and the data sets used were unpacked. The investigation of the current clinic network in the three metros made it clear that there is a need for improving the accessibility to primary health care in all three metros, especially in Tshwane. Three scenarios with differences in locations and number of households were created as test scenarios for this project. For each of these scenarios, household attribute tables are available up until 2030. The number of children, household income and distance to the nearest facility were then selected as the attributes to include in the demand calculations. A distance decay function was derived from the literature to calculate the probability of a household member visiting a clinic when ill. For the probabilities based on the household income and number of children, fixed intervals were used with probabilities assigned to the intervals. Software to assist the modeller was investigated; Python was the software selected and DEAP the package to assist with the Genetic Algorithm (GA). The model development is described in the next chapter.

Chapter 4

Model development

This chapter discusses the model developed to solve the problem and the assumptions made. To solve this problem, two models were developed. The first is a facility location model and the second is a robust facility location model. The first model was developed to determine a good configuration and variable values for each scenario. This model determines when and where to locate the container clinics from a set of available locations. In this model, three variables were minimised: the total building cost, the total travel distance from the households to the facility, and the total distribution distance.

The results of the first model were used as input to the second model. The robust facility location model also determines a good configuration and minimises the variable values. This model finds a configuration as close as possible to the individual model results for the three scenarios. The three variables minimised in the first model are also minimised in this model. However, the difference between the variable values for the different scenarios of a given configuration are minimised and the difference between the scenario variable values of the scenario and those of the selected configuration.

The assumptions made for the model and the mathematical models are described next.

4.1 Assumptions

Several assumptions were made for the model and are stated next.

- All deliveries to the clinics are made from the distribution centre identified in Chapter 3. The location of the distribution facility will have a significant impact on the output of the model since the distribution costs have an impact on the decision of where to locate the clinics.
- Only households with an annual income of R 108 00 or less are considered in this study. These are the households that are more likely to make use of public clinics and do not necessarily have sufficient household funds to travel far for medical care.
- At least 90% of the households must be within a 5 km radius of a clinic. It is not necessarily financially feasible to have primary health care within 5 km from all households. The 90% ensures that a vast majority of households have accessible health care.
- The accessibility goal has to be reached within the first five years. Since access to primary health care is a right of all South Africans, a relatively short time frame was selected to improve the current accessibility to health care.
- The cost of building a clinic in 2018 was R 171 794 (Big Box Containers, 2020).
- A yearly building cost inflation of 3.5% is applied as per the Building Cost Index (Bureau for economic research, 2019). The building cost inflation will have an impact on when the buildings are built; if the inflation is very high, it will force the model to open more clinics at the beginning to reduce the costs.
- The cost of building a clinic will be fixed, irrespective of the location. The building cost only includes the actual cost of building and not the acquisition of land and many other factors traditionally included in

the building cost. Having a fixed building cost will ensure that there is no bias against building in more expensive areas or only building in remote areas where the cost of land is cheaper.

- No operating costs are considered in the model. The model only considers the cost of opening the clinics. By not including operating costs, clinics will be built as soon as possible as only the initial construction cost is minimised.
- The distance between the facilities and zones were considered as straight line distances multiplied by a crow-fly factor of 1.265 (Barthelemy, 2011). Using this distance rather than actual road distances could mean that the facility is not within 5 km travel distance due to geographical reasons such as mountains or rivers not being part of the calculation. However, straight line distances with a crow-fly factor reduce the problem's size for not having to use a road network.

The model's health care demand was considered a dynamic variable as the distance variable used to determine the demand is a dynamic variable. All the other variables used to calculate the demand were static variables. A mathematical model was developed with these assumptions to determine a robust configuration of when and where to open clinics, given the three scenarios.

4.2 Mathematical model

This conceptual problem was converted into a more concrete problem by modelling it mathematically. The following sets and variables were defined for the model.
Let:
$\boldsymbol{E}$ be the set of existing clinics
$\boldsymbol{L}$ be the set of candidate locations for the container clinics
$\boldsymbol{S}$ be the set of scenarios, 1 = trend, 2 = economic spike, 3 relocation
$\boldsymbol{Y}$ be the set of years from 2018 to 2030
$\boldsymbol{Z}$ be the set of zones with possible patients

Let:

$$\begin{aligned}
b_y &\triangleq \text{Cost of building a clinic in year } y \in \boldsymbol{Y} \\
d_{zy} &\triangleq \text{Demand per zone } z \in \boldsymbol{Z} \text{ for year } y \in \boldsymbol{Y} \\
g_{1f} &\triangleq \text{The total cost of the good configuration for scenario } f \in \boldsymbol{S} \\
g_{2f} &\triangleq \text{The total distance of the good configuration for scenario } f \in \boldsymbol{S} \\
m_{zi} &\triangleq \text{Distance (km) between households in zone } z \in \boldsymbol{Z} \text{ and health care facility } i \in \boldsymbol{E}, \boldsymbol{L} \\
n_i &\triangleq \text{Distance (km) between the distribution centre and health care facility } i \in \boldsymbol{E}, \boldsymbol{L} \\
o_{zy} &\triangleq \text{Number of households in zone } z \in \boldsymbol{Z} \text{ for year } y \in \boldsymbol{Y} \\
p_{1zy} &\triangleq \text{Calculated number of households in zone } z \in \boldsymbol{Z} \text{ that will visit a clinic for year } y \in \boldsymbol{Y} \\
&\quad \text{ based on the probabilities} \\
p_{2z} &\triangleq \text{Probability of a household in zone } z \in \boldsymbol{Z} \text{ visiting a clinic based on the distance to the nearest facility} \\
v_{1f} &\triangleq \text{The cost of a configuration for scenario } f \in \boldsymbol{S} \\
v_{2f} &\triangleq \text{The distance of a configuration for scenario } f \in \boldsymbol{S} \\
w_{ly} &\triangleq \begin{cases} 1 & \text{if clinic } l \in \boldsymbol{L} \text{ is opened in year } y \in \boldsymbol{Y} \\ 0 & \text{otherwise} \end{cases} \\
x_{ziy} &\triangleq \begin{cases} 1 & \text{if household in zone } z \in \boldsymbol{Z} \text{ is served by clinic } i \in \boldsymbol{E}, \boldsymbol{L} \text{ in year } y \in \boldsymbol{Y} \\ 0 & \text{otherwise} \end{cases}
\end{aligned} \tag{4.1}$$

For the individual scenarios, two objective functions are minimised. The first is the total building cost calculated in Equation (4.2). The second objective function (Equation (4.3)) is to minimise the total distance travelled by the households to the nearest clinic and the total distance from the distribution centre to all the open clinics.

$$min\ z_1 = \sum_{y \in \boldsymbol{Y}} \sum_{l \in \boldsymbol{L}} b_y w_{ly} \tag{4.2}$$

$$min\ z_2 = \sum_{y \in \boldsymbol{Y}} \sum_{z \in \boldsymbol{Z}} \sum_{i \in \boldsymbol{E},\boldsymbol{L}} x_{ziy} m_{zi} + \sum_{y \in \boldsymbol{Y}} \sum_{l \in \boldsymbol{L}} \sum_{i \in \boldsymbol{E},\boldsymbol{L}} (12 - y)\, w_{yl} n_i \tag{4.3}$$

For the robust model, the two objective functions, as stated above, are minimised. The difference between the good values of the scenarios and the objective value of the current configuration is minimised in Equation (4.4). The difference between the total distance travelled for the scenarios' good solutions and the current configuration with a 90% accessibility constraint is minimised as well as the total travel distance variation of the current configuration for the three scenarios using Equation (4.5).

$$min\ z_3 = \sum_{f \in \boldsymbol{S}} |g_{1f} - v_{1f}| \tag{4.4}$$

$$min\ z_4 = \sum_{f \in \boldsymbol{S}} |g_{2f} - v_{2f}| + \sigma(v_{21},\, v_{22},\, v_{23}) \tag{4.5}$$

The model is subject to the following constraints:

$$\sum_{l \in \boldsymbol{L}} \sum_{y \in \boldsymbol{Y}} w_{ly} \geq 1 \tag{4.6}$$

$$\sum_{y \in \boldsymbol{Y}} w_{ly} \leq 1 \qquad \forall\, l \in \boldsymbol{L} \tag{4.7}$$

$$P(min(m_{zi} x_{ziy}) \leq 5) \geq 0.9 \qquad \forall\, z \in \boldsymbol{Z},\, y \in \boldsymbol{Y},\, i \in \boldsymbol{E},\boldsymbol{L} \tag{4.8}$$

$$\sum_{i \in \boldsymbol{E},\boldsymbol{L}} x_{ziy} \geq 1 \qquad \forall\, z \in \boldsymbol{Z},\, y \in \boldsymbol{Y} \tag{4.9}$$

$$\sum_{y \in \boldsymbol{Y}} x_{ziy} \leq 13 \sum_{y \in \boldsymbol{Y}} w_{ly} - \sum_{y \in \boldsymbol{Y}} y(w_{ly}) \qquad \forall\, i,\, l \in \boldsymbol{L},\, z \in \boldsymbol{Z} \tag{4.10}$$

$$p_{2z} = 0.95^{m_{zi}} \qquad \forall\, z \in \boldsymbol{Z},\, i \in \boldsymbol{E},\boldsymbol{L} \tag{4.11}$$

$$d_{zy} = o_{zy} p_{zy} p_{2z} \qquad \forall\, z \in \boldsymbol{Z},\, y \in \boldsymbol{Y} \tag{4.12}$$

$$w_{ly} \in \{0;1\} \qquad \forall\, l \in \boldsymbol{L},\, y \in \boldsymbol{Y} \tag{4.13}$$

$$x_{ziy} \in \{0;1\} \qquad \forall\, z \in \boldsymbol{Z},\, i \in \boldsymbol{E},\boldsymbol{L},\, y \in \boldsymbol{Y} \tag{4.14}$$

At least one clinic must be built; this is enforced by Equation (4.6). Equation (4.7) ensures that a clinic cannot be built more than once. Equation (4.8) ensures that there is a clinic within 5 km of 90% of the households investigated in this study. Equation (4.9) ensures that all households are serviced by at least one clinic. Equation (4.10) ensures that a clinic can only serve patients if it has been opened that year or in a previous year. The probability of a household member going to a clinic when ill based on the distance to the nearest facility is calculated in

Equation (4.11). The calculation of the health care demand based on the available household attributes and the distance to the facility is given in Equation (4.12). Equation (4.13) - Equation (4.14) are the binary constraints for the decision variables.

This mathematical model was solved using GA, which was coded in Python. The code used to solve the model is shown in the next section.

4.3 Python code

This section is a breakdown of the code and functions developed in Python to solve this problem, starting with the set-up for the GA parameters. These parameters are set using Algorithm 5. For this model, the values of the parameters are as follow. A maximum number of 50 generations was selected to reduce the running time of the model. A population size of 30 ensured a large enough population to get a wide range of solutions. The probability of crossover occurring was set to 50% and the probability of a mutation occurring was set to 40%. μ is the number of individuals to select for the next generation and λ is the number of children to reproduce at each generation. The length of an individual was set to the number of possible locations for the clinics.

Algorithm 5: Genetic algorithm parameters

```
def set_ga_parameters():
    global max_generations,hall_of_fame_size,population_size,p_crossover,p_mutation,lambda_,mu
    max_generations = 50     hall_of_fame_size = 50
    population_size = 30
    p_crossover = 0.5
    p_mutation = 0.4
    λ = int(0.6 * population_size)
    μ = int(0.6 * population_size)
    individual_length = n_open_zones
return max_generations,hall_of_fame_size,population_size,p_crossover,p_mutation,λ,μ,individual_length
```

With the parameters for the GA set, the functions to calculate the objective and variable values are developed. Algorithm 6 was used to calculate the building cost of all the clinics that are opened between 2018 and 2030. This calculation includes the building inflation of 3.5% per year. The total cost function works through the individual and if a clinic is built, it calculates the building cost for that year with b as the building cost in the first year. These building costs are added to calculate total building costs for the individual.

Algorithm 6: Total cost calculation

```
for j in range(len(individual)): do
    for i in range(2018,2031): do
        if individual[i] > 0:
            building_cost = b((1+0.035)^(i));
            total_building_cost += building_cost
        end
    end
end
return total_building_cost
```

The code in Algorithm 7 shows the calculation for the total household travel distance for all the households to the nearest clinic for all the years. This function loops through all the years to calculate the yearly and the total distance travelled. The first step is to identify all the clinics opened in the specific year or previous years and add them to the list of open clinics. This list of open clinics is used to filter the distance matrix. The filtered distance matrix is used to get the shortest distance to a clinic for all the households. These distances are stored for the yearly distance and added together to get the total household travel distance.

Algorithm 7: Total travel distance calculation

```
for q in range(1,n_years+1): do
    zone_travel_dist = 0;
    i_list = set(list(range(1,q+1)));
    indices = [z for z, y in enumerate(individual) if y in i_list] ;
    new_zone_indices = [] ;
    new_zone_indices_2 = [] ;
    for zz in indices: do
        new_zone_indices.append(df_open_zones_2.loc[df_open_zones_2['list_index'] == zz].index[0])
    end
    for zz in new_zone_indices: do
        new_zone_indices_2.append(centroids.loc[centroids['zone_id'] == zz].index[0])
    end
    keep = df.iloc[new_zone_indices_2];
    keep = keep.append(existing_df)
    for col in keep.columns: do
        if col in hh_count_df.index.values:
            n_hh = hh_count_df.loc[col,2017+q];
            keep3 = keep[col];
            acc_min = min(keep3);
            zone_travel_dist += n_hh*acc_min;
        end
    end
    dist_dict["year_0".format(q)] = zone_travel_dist;
    total_hh_travel_dist += zone_travel_dist
end
return dist_dict, total_hh_travel_dist
```

The total distance from the distribution centre to all the newly opened clinics for each year is calculated with the code in Algorithm 8. This function systematically goes through the individual and the years to determine when the clinic was opened. Once a clinic is opened, it is serviced by the distribution centre for the remainder of the period. All these distances are added to calculate the total distribution distance for the newly opened clinics.

The accessibility of the households to the open clinics was calculated using the code in Algorithm 9. The function loops through all the years and adds the newly opened clinics for that year and the previously opened clinics to the list of existing clinics. This list is then again used to filter the distance matrix. The minimum distance for the household to the clinic is determined; if it is further than 5 km, the household is added to 'not accessible'. Once all the non-accessible households have been identified, the accessibility percentage for the year is calculated. This accessibility is used in Algorithm 10 to enforce the accessibility constraint that 90% of the households must be within 5 km of a clinic after the first five years. For the given individual, the yearly accessibility is calculated as follows: If the accessibility is less than 90% after the first five years, the constraint is broken and therefore, the solution will be deemed infeasible.

Algorithm 8: Total distribution distance calculation

```
for j in range(len(individual)): do
    for i in range(n_years): do
        if individual[j] == 1:
            distribution_dist += n[j] *n_years
        end
        else if individual[j] == 2:
            distribution_dist += n[j] *(n_years-1)
        end
        else if individual[j] == 3:
            distribution_dist += n[j] *(n_years-2)
        end
        else if individual[j] == 4:
            distribution_dist += n[j] *(n_years-3)
        end
        else if individual[j] == 5:
            distribution_dist += n[j] *(n_years-4)
        end
        else if individual[j] == 6:
            distribution_dist += n[j] *(n_years-5)
        end
        else if individual[j] == 7:
            distribution_dist += n[j] *(n_years-6)
        end
        else if individual[j] == 8:
            distribution_dist += n[j] *(n_years-7)
        end
        else if individual[j] == 9:
            distribution_dist += n[j] *(n_years-8)
        end
        else if individual[j] == 10:
            distribution_dist += n[j] *(n_years-9)
        end
        else if individual[j] == 11:
            distribution_dist += n[j] *(n_years-10)
        end
        else if individual[j] == 12:
            distribution_dist += n[j]
        end
    end
end
return distribution_dist
```

Algorithm 9: Yearly accessibility calculation

```
def accessibility_func(individual):
accessibility_total = 0
d = {}
d_percentage = {}
existing_df = df.iloc[existing_clinics_zones,:]
for q in range(1,n_years+1): do
    accessibility = 0
    i_list = set(list(range(1,q+1)))
    indices = [z for z, y in enumerate(individual) if y in i_list]
    new_zone_indices = []
    new_zone_indices_2 = []
    for zz in indices: do
        new_zone_indices.append(df_open_zones_2.loc[df_open_zones_2['list_index'] == zz].index[0])
    end
    for zz in new_zone_indices: do
        new_zone_indices_2.append(centroids.loc[centroids['zone_id'] == zz].index[0])
    end
    keep = df.iloc[new_zone_indices_2]
    keep = keep.append(existing_df)
    for col in keep.columns: do
        keep3 = keep[col]
        acc_min = min(keep3)
        if acc_min > 5:
            no_accessible += 1
        end
        accessibility_total += accessibility
        d["accessibility_0".format(q)] = no_accessible
    end
    d_ percentage[q] = (100* ((n_zones-list(d.values())[q-1])/n_zones))
end
accessibility_total = sum(d.values())
accessibility_ percentage = sum(d_percentage.values()) / len(d_percentage.values())
return float(accessibility_percentage)
```

Algorithm 10: Accessibility constraint

```
def accessibility_year(individual):
    accessibility_func(individual)
    for key, value in d_percentage.items(): do
        if key > 5:
            if value < 90:
                return False
            end
        end
        return True
end
```

Algorithm 11: Evaluation function for the individual scenarios

```
def evaluate(individual):
    accessibility_percentage_1 = accessibility_func(individual)
    total_cost_1 = cost_func(individual)
    dc_travel_dist_1 = dc_travel_distance(individual)
    hh_travel_dist_1 = hh_travel_distance(individual)
    total_dist = dc_travel_dist_1+ hh_travel_dist_1
return total_cost_1,accessibility_percentage_1, dc_travel_dist_1, hh_travel_dist_1
```

Evaluation function in Algorithm 11 use all these functions to determine the variable values. Using the GA, the total cost, total household travel distance and distribution travel distance were minimised while satisfying the accessibility constraint to determine a good configuration for each scenario. The solutions obtained from the model for each scenario are unpacked in the next chapter.

The robust model uses the same functions as the individual model; the only exception is the evaluation function. The variable values minimised in the individual models are used in the goal programming function to minimise the difference between the variable value in the robust model and the scenario solutions. The standard deviation of the total household travel distances across the different scenarios with the robust configuration is minimised. The variables calculated in this goal programming function are minimised and the accessibility is calculated to ensure that the accessibility constraint is adhered to. Using Algorithm 12, a robust configuration is determined that will perform well for multiple scenarios. The robust configuration is analysed in the next chapter.

Algorithm 12: Evaluation function of the robust model

```
def goal_programming(individual):
    total_cost = cost_func(individual)
    total_cost_2_diff_base= total_cost - total_cost_base
    total_cost_2_diff_economic= total_cost - total_cost_economic
    total_cost_2_diff_reloc= total_cost - total_cost_reloc

    distribution_dist_1 = dc_travel_cost(individual)
    total_dc_dist_diff_base= distribution_dist_1 - total_dc_dist_base
    total_dc_dist_economic= distribution_dist_1 - total_dc_dist_economic
    total_dc_dist_reloc= distribution_dist_1 - total_dc_dist_reloc

    hh_travel_dist_trend = hh_travel_cost(individual,hh_count_df_trend)
    hh_travel_dist_economic = hh_travel_cost(individual,hh_count_df_economic)
    hh_travel_dist_relocation = hh_travel_cost(individual,hh_count_df_relocation)

    base_diff = abs(base_opt - hh_travel_dist_trend)
    economic_diff = abs(economic_opt - hh_travel_dist_economic)
    relocation_diff = abs(relocation_opt - hh_travel_dist_relocation)

    sd = statistics.stdev([hh_travel_dist_trend, hh_travel_dist_economic, hh_travel_dist_relocation])
    accessibility_percentage = accessibility_func(individual)
return base_diff, economic_diff, relocation_diff, cost_diff, sd, accessibiility_percentage, total_cost,
  distribution_dist_1, hh_travel_dist_trend, hh_travel_dist_economic, hh_travel_dist_relocation
```

4.4 Concluding remarks

The assumptions made in this project and the impact thereof on the model and solution are described. A few key assumptions were made that will impact the results of the model. A fixed building cost irrespective of the clinic's location is assumed, however, inflation is included in the cost. Straight line distance with a crow-fly factor of 1.265 is used when calculating accessibility. The 90% accessibility goal has to be reached within the first five years of the model and sustained from then onwards. With these assumptions, mathematical models were developed to convey the problem: a model to find good solutions for the individual scenarios and a goal programming model to find the robust solution. These models were converted into Python models to find a solution to the problem in this report with a GA. The solutions of these models are discussed in the next chapter.

Chapter 5

Solution

In this chapter, the solutions obtained from the models defined in Chapter 4 are unpacked. For each scenario, good locations for the container clinics were identified using a genetic algorithm. The locations were selected by minimising three variables: the total distances travelled by the households to their nearest clinics, the total distance from the selected distribution centre to the open clinics and the total building cost. As there is no operating cost included in the model that penalises the opening of clinics before it is required, the majority of clinics are opened in the first year to minimise the building costs. This is a known shortcoming in the model based on the assumptions made. An accessibility target of 90% was added to the model to ensure that at least 90% of the households are within 5 km of the nearest clinic within the first five years. This accessibility target requires opening a large number of clinics in the almost 10 000 km^2 area investigated. To identify robust locations for the container clinics, goal programming was used. The values of variables minimised in the scenario model were used as the goals.

The solutions for each scenario and the robust solution are described below. The placement of the clinics, the total travel distance and the improvement in accessibility are analysed.

5.1 Trend scenario solution

For the trend scenario, most of the clinics were opened in the first year to respond to the immediate demand and accessibility target that has to be reached within the first five years. Opening the facilities in the first year minimises the total building cost due to the building inflation incorporated in the model and alleviates the immediate underserved demand. Since no operating costs were included, there are no incentives to open clinics at a later stage. This skews the majority of clinics to be opened in the first year as there is no penalty for opening a clinic a few years before it is required. The rest of the clinics are opened as the demand increases over the years. The accessibility percentage and the variable values minimised by the model are provided in Table 5.1. From 2018 to 2019, there is a significant increase in accessibility from about 60% to 90%. The opening of the initial clinics also led to a large decrease in total household distance travelled as the clinics are now much closer to the households. There is also an increase in the total distribution distance from the distribution centre to all the open clinics as there are 283 more clinics that have to be serviced. From 2022, the accessibility fluctuates between 90% and 91%. This accessibility is influenced by the new households and the new clinics opened in that year. The total distance travelled by the households slowly increases per year as the number of households increase per year based on the scenario population growth and relocation rates.

The locations of opened clinics per year are indicated in red in Appendix A.1. The impact on the overall accessibility of opening these clinics is depicted in Figure 5.1 through heat maps. By opening these clinics, the accessibility to primary health care will improve significantly from the current 65% to 91% by 2030. The improvement in accessibility can clearly be seen from the base year (2018) to 2023, the first five years. After that, there are still improvements in accessibility, mainly in Tshwane, however, this change is much smaller than in the first five years.

The most noticeable improvement is from the base year to 2019 since the majority of clinics were opened in 2019. In this scenario, the accessibility changes over the years after 2019 are minimal. This is due to the fact that there is no drastic change in the population distribution and demand, given that the current growth trends are

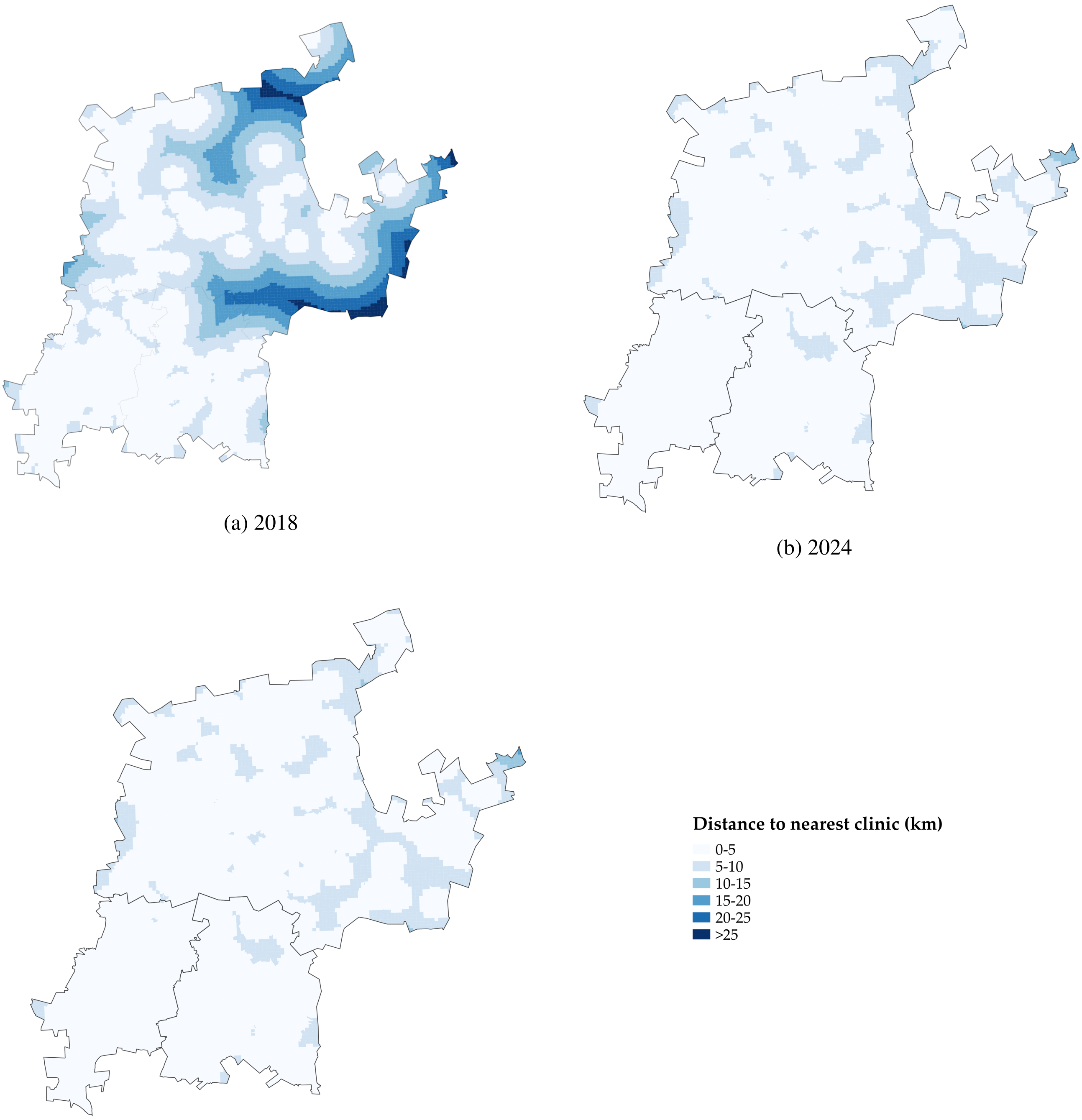
Table 5.1: Trend scenario yearly results

Year	Number of clinics built	Accessibility	Cost (R)	Total distance travelled by the households (km)	Distance from distribution center to all the opened clinics per year (km)
2018	0	59.24%	0	16 689 239	0
2019	283	90.95%	50 319 322	6 607 751	1 693
2020	1	90.95%	184 030	6 901 770	1 693
2021	0	90.95%	0	7 225 002	1 704
2022	2	90.95%	394 275	7 607 755	1 704
2023	0	90.95%	0	8 005 960	1 711
2024	1	91.08%	211 178	8 296 643	1 711
2025	0	91.08%	0	8 603 728	1 711
2026	0	91.08%	0	8 915 802	1 713
2027	1	91.13%	234 137	9 243 259	1 715
2028	1	91.15%	242 332	9 572 436	1 716
2029	1	91.28%	250 814	9 923 322	1 716
2030	0	91.28%	250 814	9 923 322	1 716

followed. The majority of clinics opened from 2020 onwards are opened in Tshwane, which leads to an improvement in the accessibility in the area. Minimal improvement can be seen in the accessibility of the households in Johannesburg and Ekurhuleni after 2019. Some clinics are opened in later years to maintain the 90% accessibility. However, there is not a drastic increase in population growth in this scenario, forcing households to relocate to the outskirts of the municipality that would require new clinics to be built.

The mean distance to a clinic is 2.8 km. This means that the clinics are accessible even without spending money on transport. A detailed breakdown of the accessibility per income class is shown in Figure 5.2. In income class 1, less than 1% of the households are between 10 km and 15 km from the nearest clinic and more than 90% within 5 km of the nearest clinics. For income class 2, just less than 90% of the households are within 5 km of a clinic and for income class 3, it is just above 90%. From the breakdown, it is concluded that health care will be much more accessible for the lowest income class. They will no longer have to pay transport fees to visit a primary health care practitioner as the average distance to a clinic is well within walking distance. According to the National Household Travel Survey conducted by Stats SA (2015), more than 20% of wages per capita of the households in the lowest income quantile are spent on public transport. By eliminating the need for transportation to health care facilities these households will have more money to spend on other essential goods and services.

When evaluating from the supplier's point of view, the average distance to a clinic from the distribution centre is 42 km. The distance from the distribution centre to the clinics follows a normal distribution, as shown in Figure 5.3. This normal distribution implies that the clinics are well spread over the area and that the clinics are indeed catering for all the communities, not just those close to the distribution centres. The supplier can use this distance distribution to evaluate the location of their facilities and whether it would make sense to open new facilities or close some facilities. In this case, the location of the facility is sufficient as it is situated practically in the middle of the three municipalities. The maximum distance from the distribution centre to a clinic is 102 km, implying that even the households on the peripheries are catered for.



(a) 2018

(b) 2024

(c) 2030

Figure 5.1: Accessibility improvement to primary health care for the trend scenario

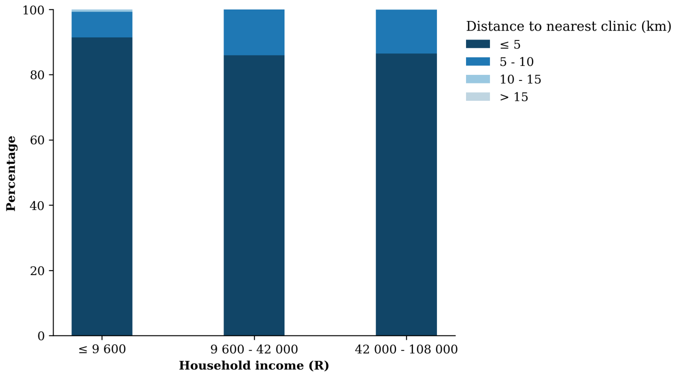


Figure 5.2: Accessibility breakdown per income category for the trend scenario by 2030

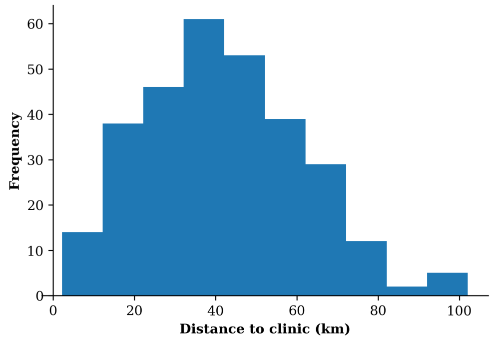


Figure 5.3: Distance distribution from the distribution centre to the clinics for the trend scenario by 2030

5.2 Economic spike scenario solution

In the economic spike scenario, the majority of clinics were opened in the first year, however, clinics were also opened in all the other years. In this scenario, fewer clinics were opened from 2019 onward than in the trend scenario. This decrease in clinics opened is due to the fact that in the economic spike scenario there is a decrease in the lower-income households in the three municipalities and, therefore, a decrease in demand for primary health care. The values of the variables minimised and the accessibility per year are summarised in Table 5.3. From 2018 to 2019, there was once again a drastic change in the accessibility and the total distance travelled by households to their nearest clinic since a lot of clinics were opened in 2019. After 5 years, the 90% accessibility target had been reached and the clinics opened from then onward was to maintain this accessibility level. The accessibility fluctuates based on the number of new households and the number of new clinics opened. The total distance travelled by households to the clinics dropped drastically in the first year as a lot of clinics were opened, reducing the distance travelled to a clinic by the households. After that the total distance travelled by the households to the clinics slowly increased per year as the number of households increased due to the population growth and household relocation. The total distance from the distribution centre to the open clinics increased as new clinics were opened.

Table 5.2: Economic spike scenario yearly results

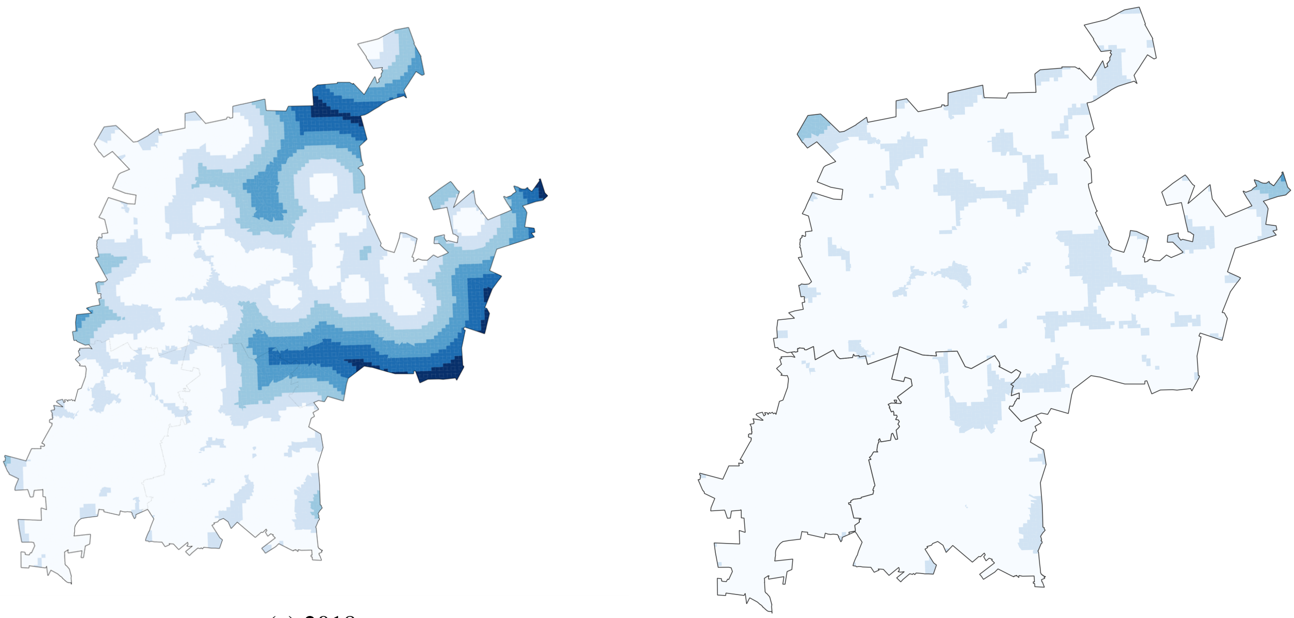
Year	Number of clinics built	Accessibility	Cost (R)	Total distance travelled by the households (km)	Distance from distribution centre to all the opened clinics (km)
2018	0	59.24%	0	16 689 239	0
2019	281	90.44%	49 963 708	10 701 071	1 713
2020	6	90.44%	1 104 180	11 129 124	1 713
2021	0	90.44%	0	11 482 942	1 717
2022	1	90.55%	197 137	11 829 994	1 735
2023	5	90.55%	1 020 187	12 215 885	1 735
2024	0	90.55%	0	12 584 012	1 743
2025	3	90.55%	655 709	12 996 574	1 747
2026	3	90.64%	678 659	13 348 939	1 750
2027	3	90.77%	702 412	13 621 233	1754
2028	3	90.77%	726 997	14 127 969	1 754
2029	0	90.81%	0	14 613 702	1 756
2030	0	90.83%	0	14 613 702	1 756

The locations of the open clinics and their impact on the accessibility through the years from 2019 to 2030 is depicted in Appendix B.1. he majority of the clinics opened from 2020 onward are located close to existing city centres because lower-income households move closer to new work opportunities. Fewer clinics had to be built in this scenario to achieve the 90% accessibility due to the reduction in the number of lower-income households.

An overview of the accessibility improvement is illustrated in Figure 5.4. By opening the clinics in the specified year and location as provided by the model, the accessibility to primary health care was drastically improved for the lower-income community while minimising the required investment. Opening these clinics led to a significant improvement in the accessibility to 90%. This accessibility is the same as for the trend scenario as once the accessibility constraint is satisfied, the model will not add additional clinics as it will only increase the cost. The model will only open new clinics if there is new demand that causes the 90% accessibility constraint to be broken.

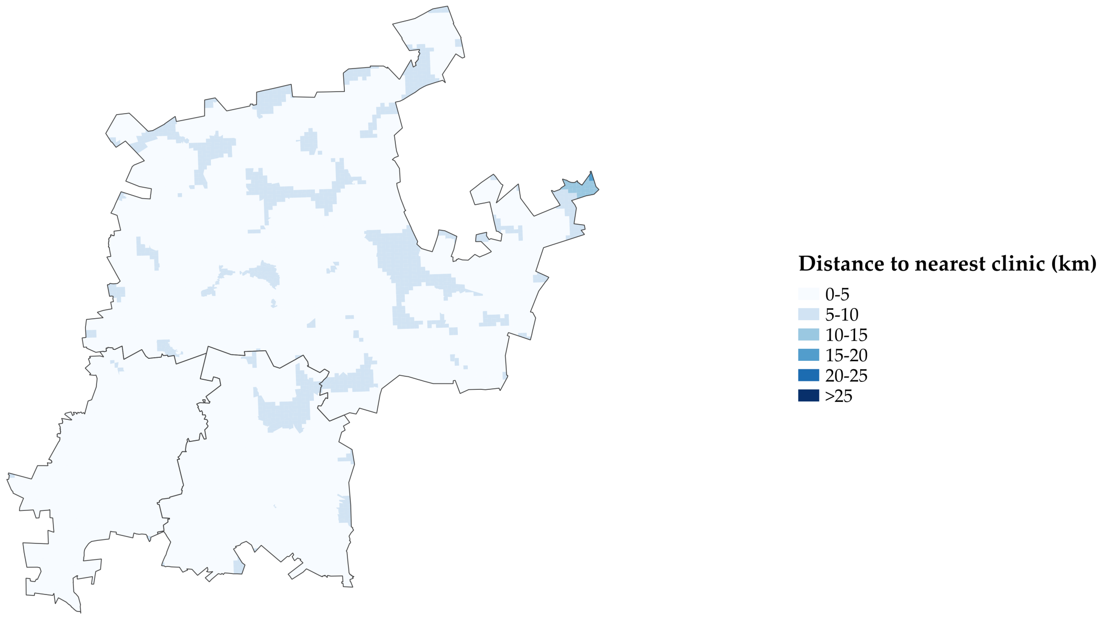
For this scenario, the mean distance to a clinic is 2.9 km, making it accessible by foot. A detailed breakdown of the accessibility per income class for this scenario is shown in Figure 5.5. Income class 1 has the highest percentage of households further than 5 km from the nearest clinic and about 2% that are further than 10 km from the nearest clinic. For income class 2, about 90% of the households are within 5 km of the nearest clinic, and for income class 3, it is almost 95%. In this scenario, the number of households in income class 1 is less than in the other two income classes and, therefore, the overall accessibility can still be above 90%. Households in these income categories will have easy access to primary health care without incurring additional costs to reach the clinic as they will be able to travel by foot.

From a supplier's point of view, this scenario is very similar to the trend scenario. The average distance from the clinics to the distribution centres are slightly lower at 41 km as more clinics are located closer to the city centres. The distance distribution once again follows a normal distribution, as shown in Figure 5.6, which confirms that the distribution centre is located centrally with regard to all the clinics. The maximum distance from the distribution centre to a clinic is 98 km. The maximum distance is not exceptionally far compared to the mean and the distance distribution follows a normal curve. Therefore, there will not be a need to open a second distribution centre closer to some of the clinics.



(a) 2018

(b) 2024



(c) 2030

Figure 5.4: Accessibility improvement to primary health care for the economic spike scenario

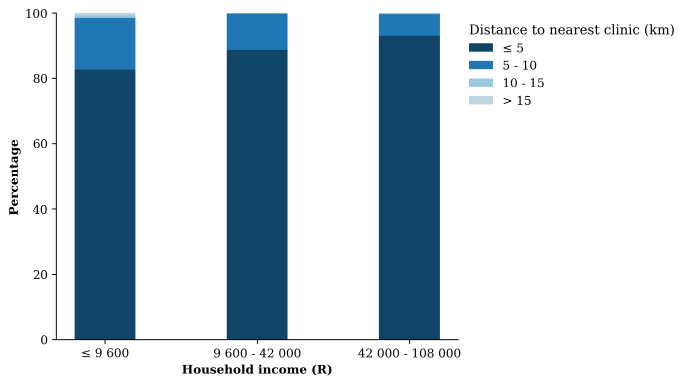


Figure 5.5: Accessibility breakdown per income category for the economic spike scenario by 2030

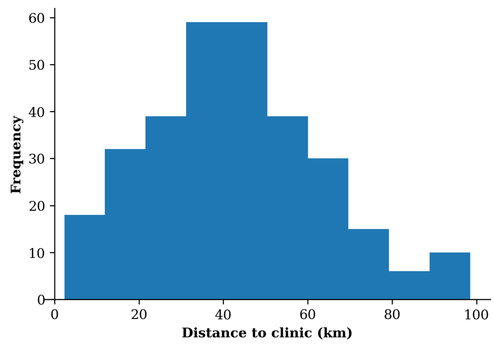


Figure 5.6: Distance distribution from the distribution centre to the clinics for the economic spike scenario by 2030

5.3 Relocation scenario solution

Most of the clinics for this scenario were once again built in the first year as there was an immediate need for clinics and it was the cheapest year to build the needed clinics. The rest were built as the demand created a need. The clinics built in the first five years are located more towards the periphery of the municipalities due to the strong spatial expansion focus to urban growth included in this scenario. The variables calculated and minimised in the model are shown per year in Table 5.2. The total cost per year was the highest in 2019 as most of the clinics were built in that year. The total distance travelled by households increased over the years even though the accessibility is relatively constant, around 90%. This is a steady increase over the years that is linked to the population and urban growth. The total distance travelled from the distribution centre to the open clinics increased each year. This increase is expected as the total number of clinics serviced each year increased. In this scenario, the total distance travelled by the households is noticeably higher than in the other two scenarios. This noticeable difference can mainly be attributed to the fact that there are much more households in this scenario than in the other two.

Table 5.3: Relocation scenario yearly results

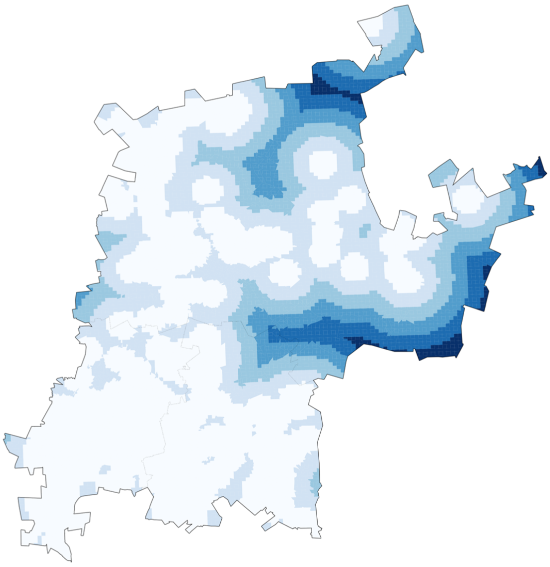
Year	Number of clinics built	Accessibility	Cost (R)	Total distance travelled by the households (km)	Distance from distribution center to all the opened clinics (km)
2018	0	59.24%	0	16 689 239	0
2019	226	88.21%	40 184 335	7 690 439	1 359
2020	8	89.54%	1 472 240	7 967 687	1 428
2021	11	89.54%	2 095 182	8 337 128	1 432
2022	1	90.50%	197 138	8 629 444	1 486
2023	14	90.80%	2 856 523	9 026 803	1 512
2024	6	91.00%	1 267 072	9 279 626	1 532
2025	9	91.00%	1 967 129	9 586 727	1 539
2026	3	91.35%	678 660	9 876 771	1 554
2027	7	90.11%	1 638 963	10 030 911	1 564
2028	8	90.30%	1 938 659	10 074 201	1 572
2029	7	90.50%	1 755 698	10 307 599	1 579
2030	0	90.50%	0	10 307 599	1 579

The population growth due to the relocation in this scenario led to some households being forced to move to the outskirts of the municipalities. These were the lower-income households as they could no longer afford real estate in the city centres. The location of all the clinics and the impact of opening these clinics for this scenario is depicted in Appendix C.1. In this scenario, more clinics were built after 2019 than in the other two scenarios as there was greater population growth and, therefore, greater demand. After the first five years, more clinics were opened on the municipalities' outskirts to cater to the overflow in the city centres, where in the first five years, more clinics were opened closer to the city centres to cater to the existing demand points. Even though it was the cheapest option to open all the clinics in the first year as no operating costs were considered, the trade-off with the total distribution distance forced the model to open clinics in later years. The model added the distance from the distribution centre to the clinic for all the years that it had been open. The clinics opened after the first five years were on peripheries of the municipalities and were much further from the distribution centre than the clinics closer to the centre of the municipalities. The total distribution distances were minimised by the model and these unnecessary distances accumulated forced the model to open clinics in later years.

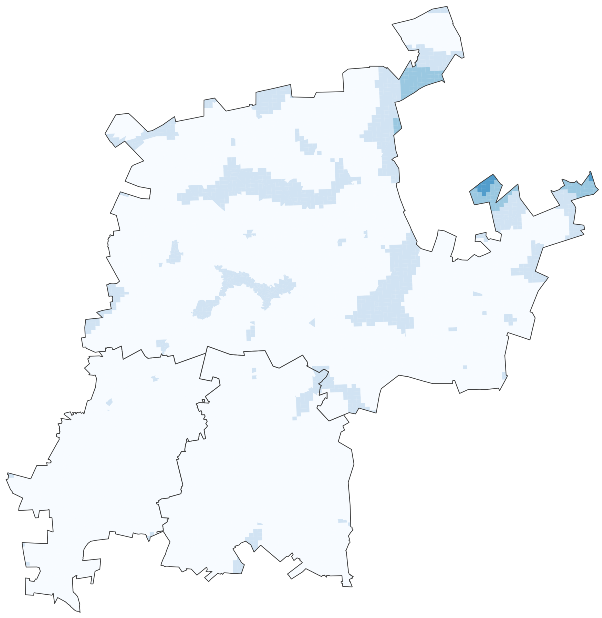
The impact of opening these clinics on the overall accessibility over the years is illustrated in Figure 5.7. Once a clinic is opened, the area around it becomes a lighter shade of blue. More clinics were required to achieve the desired accessibility since there is a greater demand due to the rapid population growth. The clinics opened after 2019 are spread across the three metropolises and not primarily in Tshwane as in the other two scenarios. The rapid population growth created new demand everywhere and not only in the already dense areas.

The mean distance to a clinic is 2.8 km. This implies that the clinics are accessible even without spending money on transport. For 90.8% of the investigated population, there is a clinic within 5 km from their homes by 2030. A breakdown of accessibility per income class is shown in Figure 5.8. Income class 2 has the highest percentage of households further than 5 km from the nearest clinic, with about 15%. For the other income classes, this is only about 10%. Even though some households are further than 5 km from the nearest clinic, it is still accessible for pedestrians. Having a clinic within 5 km makes health care accessible without incurring additional travelling costs for those who cannot afford it. City planners and decision makers can use the model to plan the budget and human resources required to realise their accessibility goal for this given scenario.

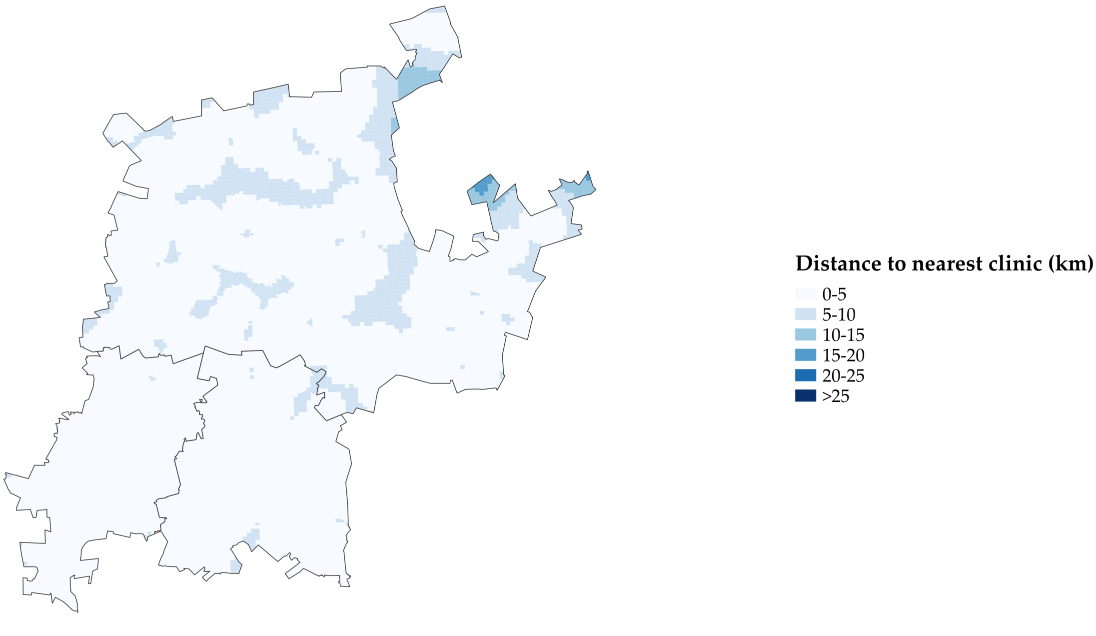
Even though the locations of the clinics opened in this scenario are more widespread, the distribution distance also follows a normal distribution, as shown in Figure 5.9. Distribution companies can use this distance distribution to determine if the current distribution network is sufficient for future development. The distribution centre is centrally located to all the clinics that it serves.



(a) 2018



(b) 2024



(c) 2030

Figure 5.7: Accessibility improvement to primary health care for the relocation scenario

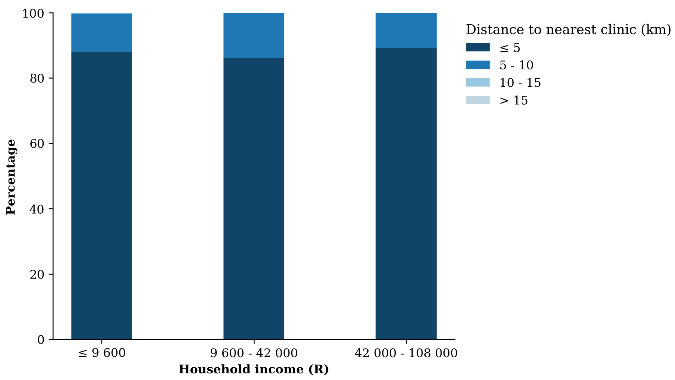


Figure 5.8: Accessibility breakdown per income category for the relocation scenario by 2030

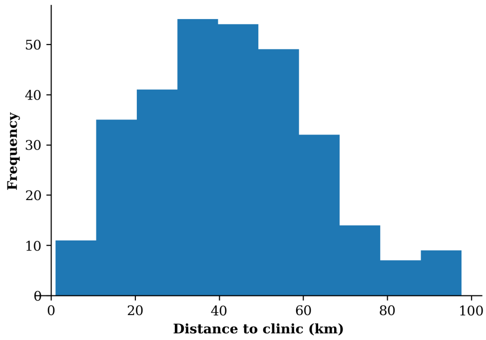


Figure 5.9: Distance distribution from the distribution centre to the clinics for the relocation scenario by 2030

5.4 Scenario comparison

Each scenario has a specific configuration of when and where to open clinics. When comparing the configurations, only 57 of the almost 300 clinics were opened in the same location and of those, 45 were opened in the same year. The location of the 45 clinics with the same configuration across all the scenarios is illustrated in blue in Figure 5.10. These clinics are in areas far away from an existing clinic and therefore high levels of unserved demand. Between the trend scenario and the economic spike scenario, only 45 clinics were opened in the same location and year. The trend and relocation scenarios have 51 clinics with the same configuration. The economic spike and relocation scenario have an exact configuration match for 52 clinics. Almost 80% of the clinics are in completely different locations. Therefore, a good solution for one scenario is not necessarily a good solution for another scenario.

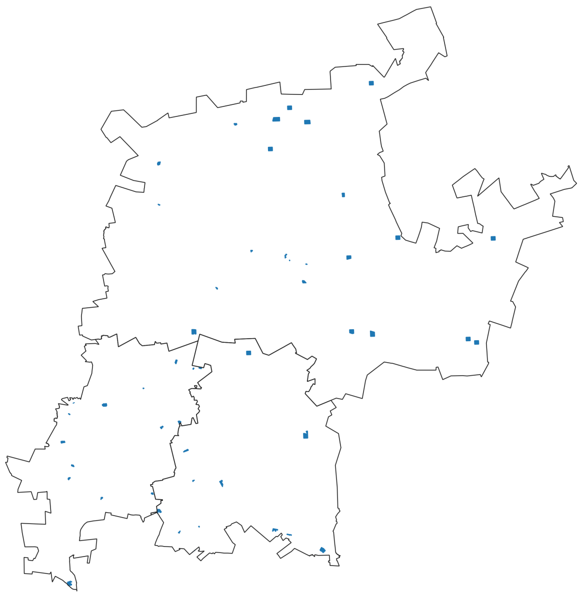


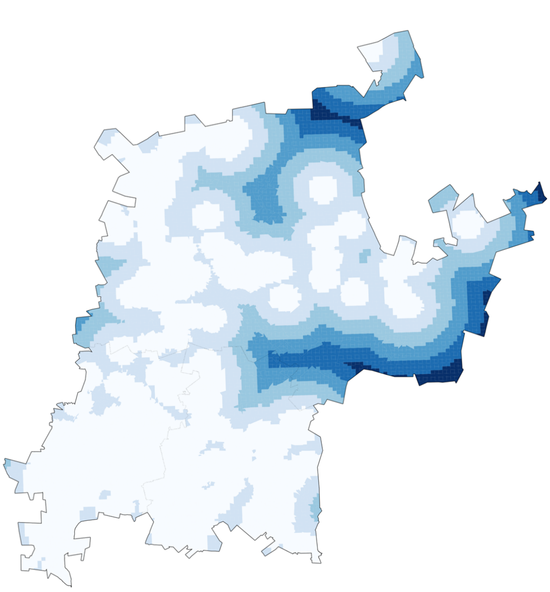
Figure 5.10: Clinic locations with the same configuration for all three scenarios

If the decision makers are confident which specific scenario will play out, they can optimise the most likely scenario and base their decisions on the results thereof. Using the scenario solutions will work well for the given scenario and the most cost-effective configuration can be determined while reaching the desired level of accessibility. The model can be run with different accessibility or cost constraints to create a pareto frontier of the accessibility, total cost, total or average travelling distance for the households to the clinics, and the total distribution distance. The pareto frontier will allow the planners to compare the trade-off between the selected objectives and therefore provide an evidence-based decision tool for a specific scenario. The frontier can also assist in determining the accessibility cut-off, when the additional costs incurred will have very little increase on the accessibility. This tool can be used for strategic decision-making support to estimate the budget and human resources required in the next 10 to 20 years. However, if, as in many real-world situations, multiple scenarios could play out, the robust model will provide the decision makers with an evidence-based decision tool that caters for the different scenarios and not just the one.

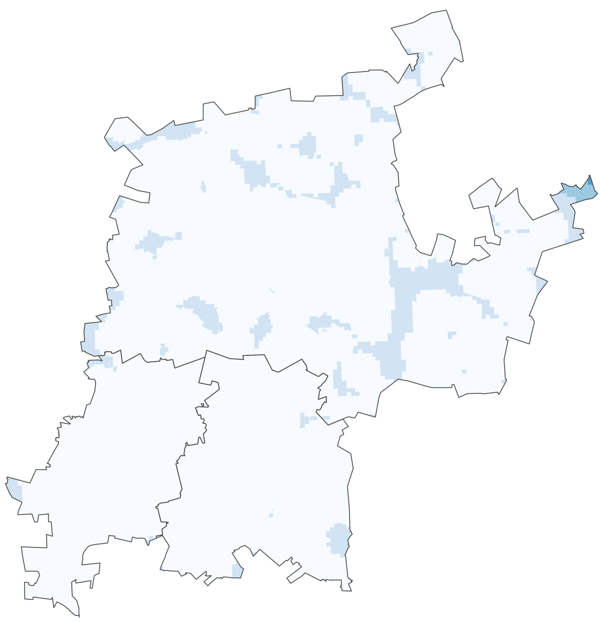
5.5 Robust solution

The robust model used goal programming to find an acceptable solution for all three scenarios while staying as close as possible to the individual scenario solutions. The objectives of the individual scenarios were set as the goals in the goal programming model to find a robust configuration. The robust model placed facilities over the years in order to come as close to the individual scenarios as possible. It seeks the best compromise between the three scenarios. For the robust scenario, the number of clinics opened is more than in the trend and economic spike scenarios as there is a greater demand base in the relocation scenario. The locations and the year in which the clinics were opened are shown in Appendix D.1. Once again, most of the clinics were opened in the first year to meet the immediate need in all the scenarios. The clinics opened after the first five years is spread across the municipalities and outskirts. A summary of the accessibility improvement is provided in Figure 5.11. Fewer clinics were opened on the outskirts that cater for the relocation scenario demand, however, enough clinics are still in operation that can cater for the demand, should it arise.

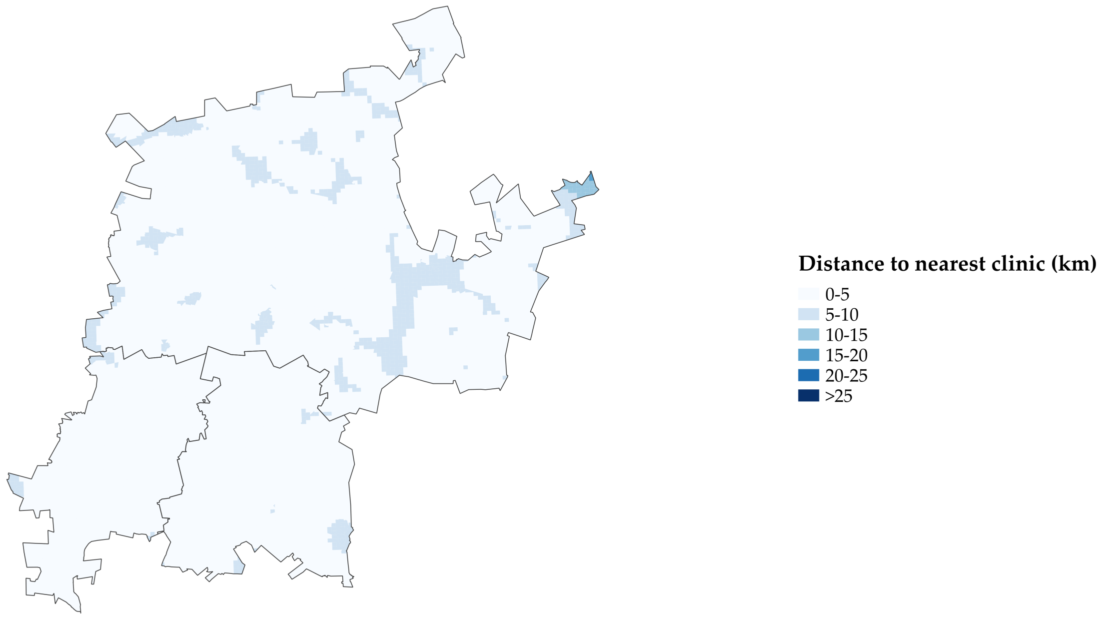
Since the majority of clinics were opened in the first year to minimise the cost and adhere to the accessibility constraint, the most significant improvement was in 2019 from the base year. With these locations of the clinics, a detailed breakdown of accessibility per income class per scenario can be seen in Figure 5.12. For all the scenarios, the percentage of households within 5 km of a clinic given the robust configuration is above 90%. Therefore, it shows that the robust configuration adheres to the accessibility constraint in all the scenarios. No households are further than 10 km from a clinic in any of the scenarios given the robust configuration. The economic spike scenario has the worst accessibility of the three scenarios given the robust configuration, but it is still within the defined constraints. The economic spike scenario has the worst accessibility because the trend and relocation scenarios add greater weight to the locations on the outskirts.



(a) 2018

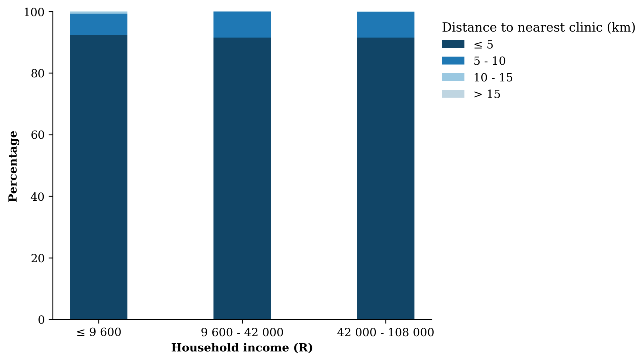


(b) 2024

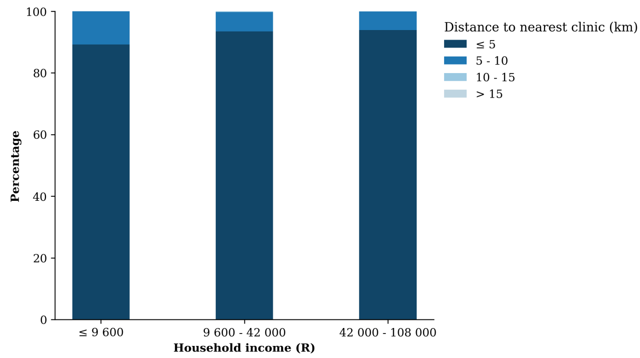


(c) 2030

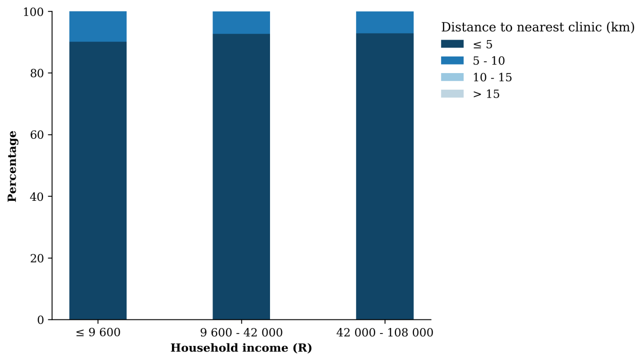
Figure 5.11: Accessibility improvement to primary health care for the robust scenario



(a) Trend scenario



(b) Economic spike scenario



(c) Relocation scenario

Figure 5.12: Accessibility breakdown per income category for the three scenarios given the robust solution by 2030

5.5.1 Objective values comparison

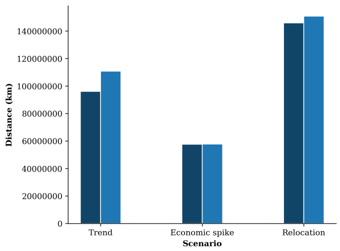
A comparison of the scenario solutions and the robust solution in each scenario is provided in Figure 5.13. The total distance travelled for the households in each individual scenario is less than the total distance for the robust configuration. This is expected since the robust configuration is not the optimal configuration for any of the scenarios, but it is a better overall solution. The robust model tries to find an acceptable compromise between all scenario; therefore, it will not be optimal for any of the scenarios but as close to scenario solutions as possible for all the scenarios.

The total distance travelled by the households for the scenarios, given the robust configuration, is similar for the trend and the relocation scenario. However, for the economic spike scenario, it is much less. This difference in household travelling distance is due to the economic spike scenario having higher densities of people located close to city centres. Therefore, more people have shorter distances to travel to clinics. When considering the robust solution, there is a much greater difference between the relocation scenario and robust total household travel distance than in the other two scenarios. This large difference is because the robust scenario does not specifically cater for all the new households located on the outskirts of the municipalities as the relocation scenario does. If a clinic is not opened within 5 km of a household, the households on the outskirts of the municipalities will have a further distance to travel to the nearest clinics than in the city centres. Therefore, the total distance of the relocation scenario quickly increases as the average total distance travelled to the nearest clinic for the robust relocation scenario is larger than in the other scenarios. For the trend and the economic spike scenarios, the total distance travelled by households for the robust configuration is close to the scenario solutions. This smaller difference can be attributed to the fact that more households are located in their original locations, close to the city centres, where the majority of the clinics are opened before any growth happens and, therefore, this ensures shorter travel distances for the households. There is almost 1% difference between the economic spike scenario solution and the robust configuration. The relocation also has a small difference of about 4%. The trend scenario has the highest difference between the scenario configuration and robust configuration with almost 14%.

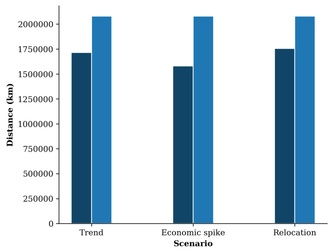
The total distribution distance of the three scenarios are all relatively close to each other. The small change can be due to the fact that the distribution centre is located more or less in the centre between the three municipalities. The total distance travelled from the distribution centre to the open clinics for the robust configuration is slightly higher than the scenario solutions. The difference between the individual scenario solutions and the robust solution are between 15% and 24%. The economic spike scenario has the biggest difference between the scenario solutions and robust total distribution distance since in this scenario, there are not many clinics placed in the outskirts of the municipalities. Most of the clinics are located close to the city centres and are closer to the distribution centre. More clinics are located on the edges of the municipalities in the robust configuration, leading to an increase in total distribution distance. As more clinics are opened in the robust solution than in the scenarios to cater for the uncertain demand, the total distribution distance for the other two scenarios is also higher than with their individual configurations. Distribution companies can use these differences in distance to determine whether or not the distribution centre needs to be relocated or if a new distribution centre should be added. The current data shows no significant need for an additional distribution centre as the distribution distance between the scenarios is fairly similar and the difference between the scenario and robust distances is not significant.

The total costs for the individual scenarios are very similar. This is because the majority of the clinics were opened in the first year and, therefore, inflation had a small effect on the total cost. The total cost for the robust configuration is higher than the scenario total cost for each scenario, as seen in Figure 5.13c. This cost difference can be attributed to the fact that more clinics were opened in the robust scenario than in the individual scenarios to deal with the uncertainty. The difference in the total cost between the scenarios and for the robust configuration is due the years in which the clinics were opened. The percentage change in the total cost between the scenario solution to the robust solution for the investigated scenarios are very similar, 18% for the trend scenario, 14% for the economic spike scenario and 16% for the relocation scenario. These differences are not really significant when dealing with a robust solution for three totally different development and growth scenarios.

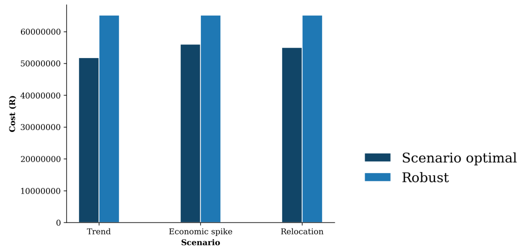
For all the variables, the difference between the robust configuration and the scenario configuration is within a 25% range. Thus, the robust configuration will perform well or at an acceptable level in all the scenarios. In this case, if all three variables are considered, the solution is a good enough solution to combat the uncertainty of how the metropolises will develop in the future. City planners can look specifically at the total household travel distance and the total cost to determine the feasibility of the robust solution, while pharmaceutical or logistics companies can do their strategic planning using the total distribution distance.



(a) Total household travel distance to the allocated opened clinics per scenario



(b) Total travel distance from the distribution centre to the open clinics per scenario



(c) Total cost to build the opened clinics per scenario

Figure 5.13: Comparison of the scenario solutions and the robust solutions per scenario

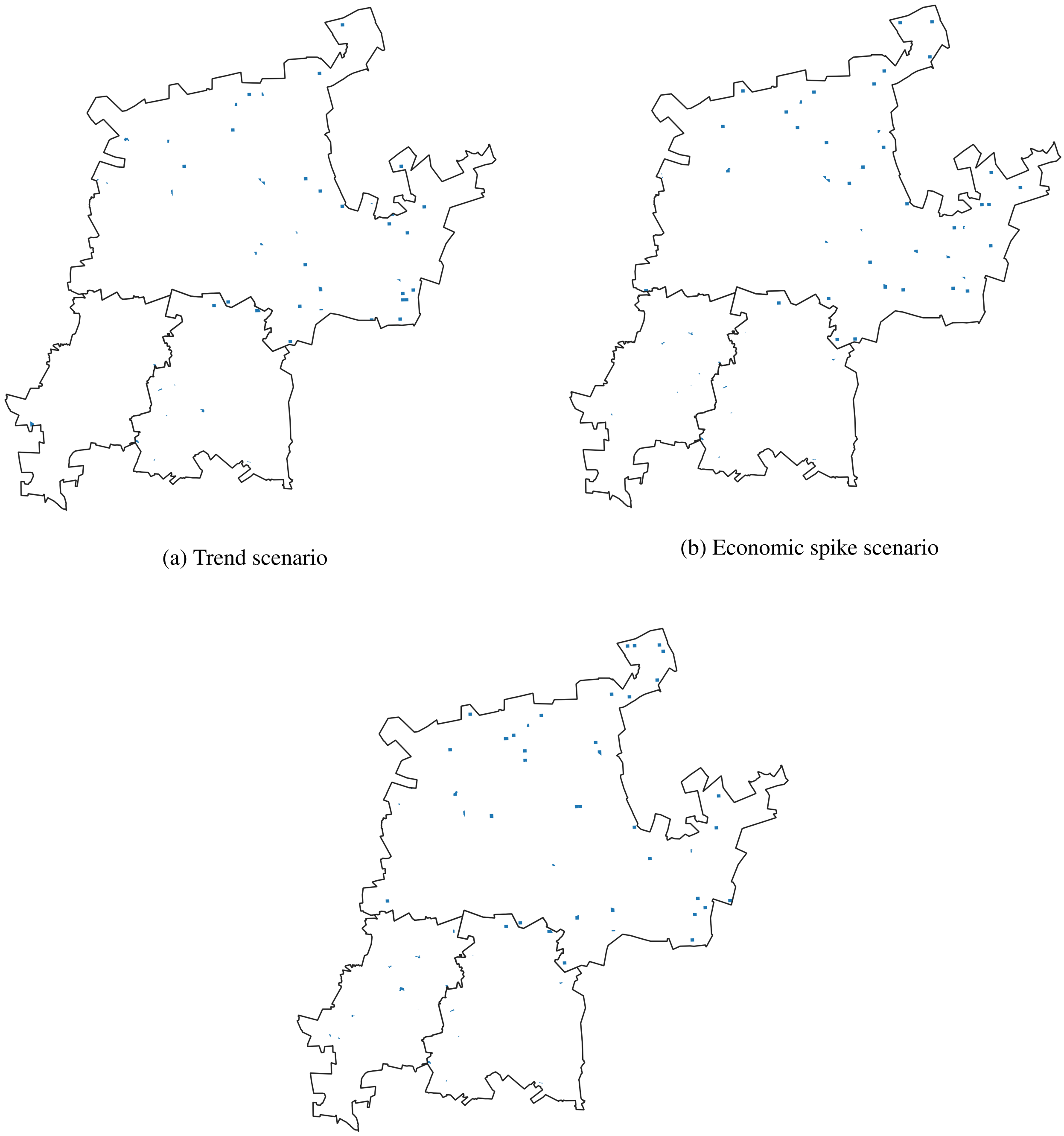
5.5.2 Configuration overlap

This robust solution has a 36% configuration overlap, where clinics were opened in the same location and year, with the trend scenario. For the economic spike scenario, the configuration overlap and the zone overlap are the same with a 36% overlap. Lastly, for the relocation scenario, almost 40% of the clinics have the same configuration. The robust solution has an almost 40% configuration overlap with all the individual scenarios. The similarities between the robust configuration and the scenario configurations are much higher than the similarities between only the scenario configurations. Therefore, it can be concluded that the robust configuration is a good compromise between the different scenarios.

The zones that have the exact same configuration in the robust solution and the scenario solutions are mapped in Figure 5.14. The majority of the overlapping is in the Tshwane municipality; this is the municipality that had the lowest accessibility to clinics. The location of the overlapping zones shows that the model caters for this lack of accessibility. The overlapping configuration in the Tshwane municipality is expected as there is a large area with low accessibility and, therefore, the clinics open in the same place every time to cater to this lack in initial accessibility. These overlapping zones should be noted as they are the zones that will definitely improve accessibility. Decision makers can use these overlaps to prioritise the clinics to be built when there are time, budget, or other constraints.

The variables investigated in this study and others can be used to provide quantitative decision support for the Department of Health and other key role players when deciding where and when to open new clinics. Trade-offs can be made between the accessibility for the communities and the cost of opening new clinics from a health care provider perspective or between the average or total travel distance and the total cost of opening the clinics. These trade-offs can be done by means of a pareto frontier. The pareto frontier will assist the decision maker to choose a solution most suitable to their current strategic needs. The pareto frontier provides a visualisation of the impact of changing the accessibility on all the other variables. The decision maker can easily see the magnitude of the impact and make trade-offs between the accessibility percentages and the variable values to ultimately determine a feasible accessibility strategy. The model provides a strategic plan of when and where to open new clinics given an accessibility goal and uncertainty of the future demand. It allows the decision makers to plan the budget and human resources for the next five to 15 years to achieve the desired health care-related goals.

Opening most of the clinics in the first year is not a realistic representation of reality, therefore, a budget analysis was conducted to test the model in a scenario closer resembling reality. In the next chapter, an accessibility analysis and a budget analysis are conducted to ensure that the model is not only sufficient for these scenarios but also for other scenarios and changes.



(a) Trend scenario

(b) Economic spike scenario

(c) Relocation scenario

Figure 5.14: The zones with the exact same configuration in the robust solution and scenario solutions

5.6 Concluding remarks

In this chapter, the solutions to the individual scenarios and the robust scenario were analysed. All the results are skewed to locate as many as possible clinics in the first year when it is the cheapest as there is no penalty in the form of operating costs added to the model. A good configuration for one scenario is nowhere near the good configuration for the other scenarios with a maximum overlap of just under 20%. The robust solution has a 40% overlap with each of the individual scenarios, which is a significant improvement. This improvement comes at a cost; the total cost of the robust scenario is higher than all three of the individual scenarios. The total household distance travelled is very similar in the trend scenario and the economic spike scenario, however, for the relocation scenario it is much higher. This large difference in the relocation scenario can be worrisome if it is a highly plausible scenario. These models can be used as tools by decision makers during strategic planning to make informed decisions based on quantitative analyses of where and when to open clinics that will provide the desired accessibility in any of the investigated scenarios.

Chapter 6

Accessibility and budget analysis

Additional analyses were conducted to investigate the impact of different accessibility percentages and the implications of modelling a more realistic scenario on model results. The first analysis looks at the impact of changing the accessibility level on the variables minimised in the model. The second analysis introduces a budget constraint to get a more realistic placement of clinics.

6.1 Accessibility analysis

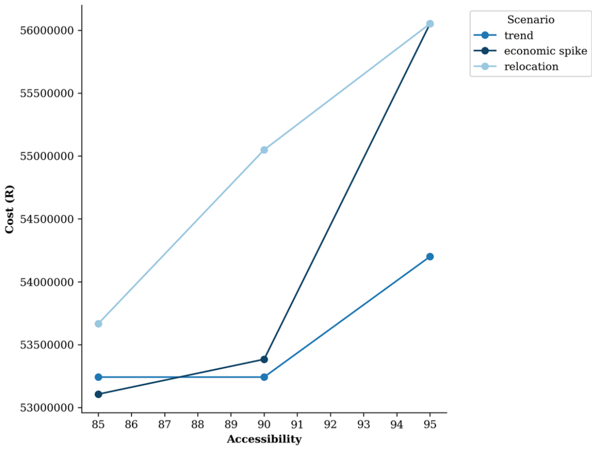
Three accessibility percentages were selected to determine the impact of accessibility on all the other variables. The accessibility percentages used were 85%, 90% and 95%. The individual scenario model was run for all the scenarios and accessibility percentages by only changing the accessibility constraint from 85% to 90% and, finally, to 95% in the model. Once the scenario solutions and accessibility were determined, the variable values were put into the robust model as the goals for each accessibility percentage where the accessibility constraint was once again adjusted to the imposed accessibility percentage.

Ignoring operating cost in the model skewed the results to open most of the clinics in the first year to minimise the total cost. However, minimising the total distribution distance ensured that some of the clinics, especially on the outskirts, were opened in later years.

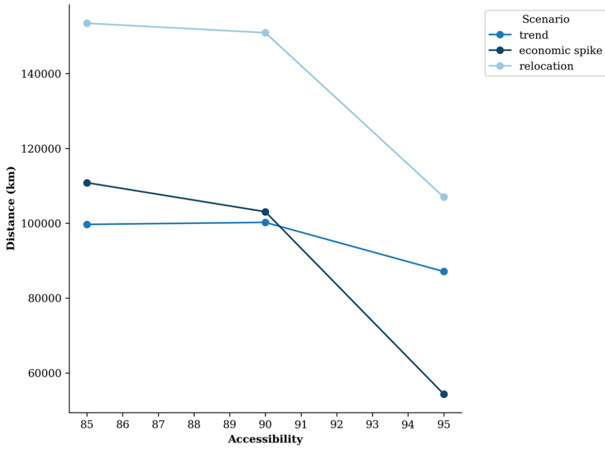
For the individual scenarios, all the variable values changed as expected. The results per scenario are shown in Figure 6.1. The total cost for the relocation scenario is the highest as this scenario requires the most clinics irrespective of the accessibility due to the number of households in this scenario. The total cost increases as the accessibility percentage increases as more clinics are built to satisfy the higher accessibility constraint. The increase in total cost for the relocation scenario from 85% to 90% and from 90% to 95% have similar gradients. However, for the trend and economic spike scenarios, the jump from the 90% to 95% is much higher than from 85% to 90%. When looking at the scenarios individually, the benefit of the additional cost to get the 95% accessibility will have to be traded off with the benefits gained. For both the trend and the economic spike scenario, the additional cost of moving from 85% accessibility to 90% is relatively small, however, for the relocation scenario, it is a more considerable increase. Moving from 85% accessibility to 90% accessibility would be advisable when only looking at the costs as it is only a slight increase in the total cost. The impact on the other variables will have to be investigated further to make well-informed decisions on which accessibility percentage is the most beneficial.

With more clinics being built as the accessibility level is increased, the total distance travelled by households to the nearest clinic reduces as more households have shorter distances to travel to the nearest clinic. Once again, the relocation scenario has the longest total distance travelled by households mainly because this scenario has much more households than the other two scenarios. Moving from 85% accessibility to 90% accessibility has little impact on the total distance travelled for all three scenarios. However, when increasing the accessibility level to 95%, there is a significant decrease in the total distance travelled by the households, especially in the economic spike scenario and the relocation scenario. This drastic change is because the model places clinics in the more dense areas first, where a larger group can be served, leaving the less dense communities further than 5 km from clinics, which can significantly increase the total distance travelled by households. When moving to 95% accessibility, the model is forced to locate clinics in these communities, which reduces the number of households that are more than 10 km from the nearest clinics. Therefore, a sharp decrease in the total distance travelled by households can be seen.

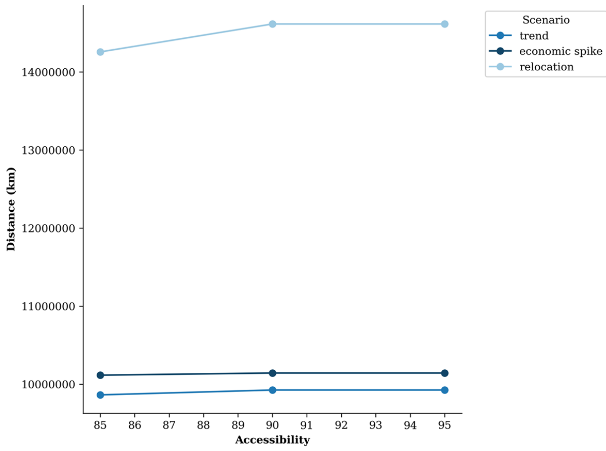
The total distance travelled from the distribution centre to the clinics increases as the accessibility constraint is tightened; this increase is mainly due to more clinics that have to be serviced. The distance travelled by households to the nearest clinics is much higher in the relocation scenario than in the other scenarios. The main reason for this large difference is the number of households, leading to more clinics that have to be opened, and the relocation scenario has much more households in the investigated income classes than the other two scenarios.



(a) Total cost to build the clinics to be opened



(b) Total household travel distance to the nearest clinic



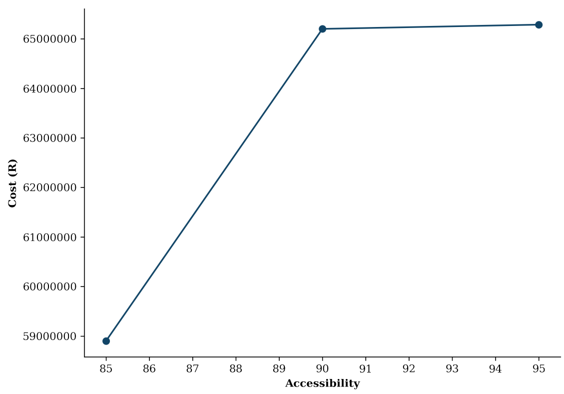
(c) Total travel distance from the distribution centre to the open clinic

Figure 6.1: The variable values of the individual scenario solutions given different accessibility percentages

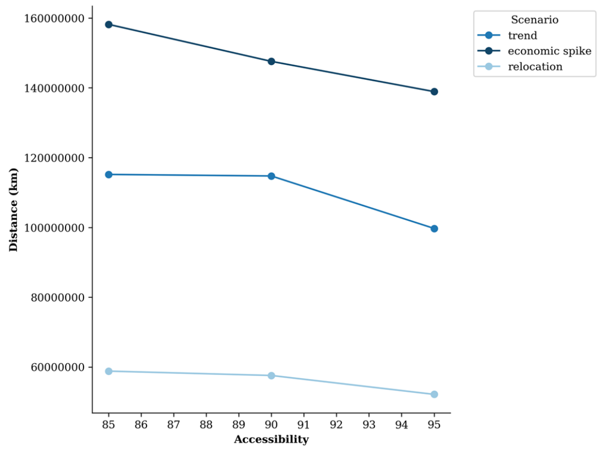
When comparing the robust solutions of these different accessibility percentages, the change seen in the individual scenarios is still present. The graphs in Figure 6.2 illustrate the impact of changing the accessibility target on the variables for the robust solutions. To get from 85% to 90% requires a large jump in cost. The cost difference between 90% accessibility and 95% accessibility is much smaller. Therefore, many more clinics are required to move from 85% to 90% accessibility than from 90% to 95%. The feasibility of the large cost difference from 85% accessibility to 90% accessibility will have to be weighed up against the benefits gained for the community members. The distance to the nearest clinic for the households further than 5 km has to be analysed to determine the additional benefit to the community. If the households are further than 10 km from the nearest clinic, the investment might be worthwhile. However, if the households are less than 10 km from the nearest clinic, the large investment required will not necessarily have the same benefits for the community as the clinics are still relatively accessible by foot.

The total distance travelled by households reduces as the accessibility target increases. This decrease is expected as a smaller portion of the households have to travel more than 5 km to the nearest clinic. The relocation scenario has the longest total distance travelled as it has the most households, followed by the trend scenario and then lastly, the economic spike scenario, which has the least number of households. The benefit of moving from 85% accessibility to 90% is relatively small in all three scenarios and would not justify the significant cost increase. There is a greater impact on the total distance travelled by households when moving from 90% accessibility to 95% accessibility. This improvement will come at little cost if the accessibility is already at 90%. If the accessibility is less than 90%, this reduction in the distance travelled by households would come at a much higher cost. The total distance from the distribution centre to the clinics increases as more clinics are opened that have to be serviced as the accessibility target increase. Moving from 85% accessibility to 90% accessibility leads to a steep increase in the total distribution distance. This increase in the distribution distance can impact the prices paid by the patients for medication as the distribution costs are ultimately passed on to them. This large increase in distance can be attributed to the fact that more clinics are opened on the outskirts of the municipalities where the population is less dense. These clinics on the outskirts are much further away from the distribution centre than clinics close to city centres and, therefore, significantly impact the total distribution distance. Improving the accessibility to 95% has a smaller impact on the total distribution distance as fewer clinics have to be opened to reach it when looking at the total cost.

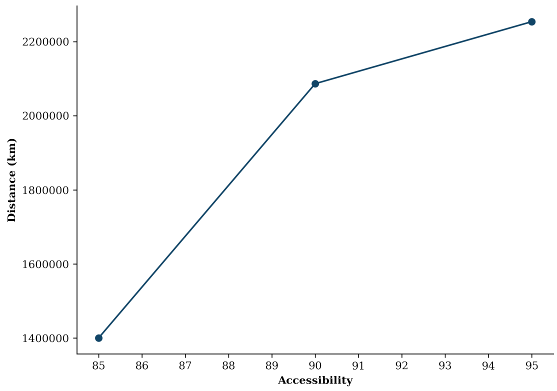
These graphs can be used as decision support to decide which accessibility target is feasible given the total cost, household travel distance and distribution distance trade-offs. The benefits of each accessibility target can be weighed up against the cost and impact.



(a) Total cost to build the clinics to be opened for the accessibility robust configuration



(b) Total household travel distance to the nearest clinic for the accessibility robust configuration



(c) Total travel distance from the distribution center to the open clinic for the accessibility robust configuration

Figure 6.2: The variable values of the robust solutions given different accessibility percentages

6.2 Budget analysis

For the budget analysis, a budget constraint was added to the model to test if the model is responsive to a restrictive budget, limiting the number of clinics that can be built per year. A random generator was used to set a budget per year. The budget that was used is provided in Table 6.1. An additional constraint was added to enforce the budget where the total cost for the year has to be less than or equal to the budget for that year.

Table 6.1: Yearly budget

Year	Budget (R)
2019	10 000 000
2020	11 400 000
2021	12 898 000
2022	14 500 860
2023	16 215 902
2024	18 051 003
2025	16 102 006
2026	15 014 600
2027	15 014 600
2028	12 115 620
2029	14 363 701
2030	16 706 918

As there is a constrained budget, the majority of clinics could not be opened in 2019, as in the original case, therefore the year by which the accessibility goal has to be reached had to be changed to allow a gradual improvement in the accessibility. In the model, the accessibility constraint was relaxed to be reached by the end of 2030 and no longer after the first five years as in the original case. By adding the budget constraint, the impact of the omitted operating costs is less significant, forcing the model to open the clinics more gradually. However, some clinics may still be opened sooner than necessary if the budget allows it and it is not far from the distribution centre, keeping the distribution distance as small as possible.

The configuration for each scenario with the adjusted parameters was determined by running the model with the budget constraint and the relaxed accessibility constraint. By adding a constrained budget, the facilities were no longer primarily opened in the first year; the placement of facilities are spread more evenly across all 12 years.

The total cost per year for the three scenarios is provided in Table 6.2. Spreading the opening of clinics over all the years leads to an increase in the total cost due to inflation, however, it is more feasible to open 50 clinics per year than almost 300 in one year. In this more realistic scenario, the relocation scenario still has the most clinics opened to cater for the fast-growing demand, followed by the trend scenario. The economic spike scenario has the least number of clinics opened as this scenario has a much slower lower-income population growth rate than the other two scenarios. When comparing the actual costs against the budget, such a large budget is not necessary. A maximum budget of about R 12 000 000 should be sufficient for all the years.

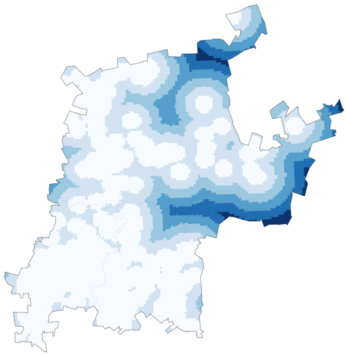
Table 6.2: Total cost per year of opening the clinics for the three scenarios

Year	Trend scenario total cost (R)	Economic spike scenario total cost (R)	Relocation scenario total cost (R)
2018	0	0	0
2019	9 601 567	8 890 339	9 423 760
2020	8 833 441	7 913 291	6 072 991
2021	8 380 727	4 952 248	8 761 670
2022	9 068 328	10 251 153	9 068 328
2023	8 569 570	9 181 682	7 549 383
2024	6 546 539	6 968 897	6 968 897
2025	5 027 109	8 742 798	7 649 948
2026	8 596 356	7 012 817	11 763 434
2027	8 428 953	7 258 265	11 472 742
2028	10 420 293	7 512 305	702 7640
2029	10 283 376	10 534 190	12 039 074
2030	0	0	0

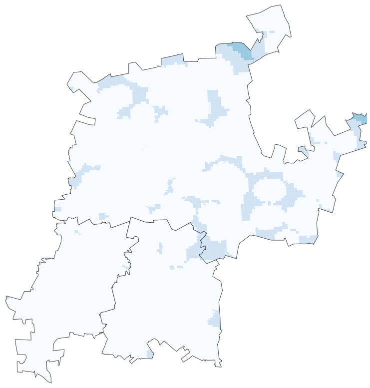
Both the trend scenario and the economic spike scenario see a large initial decrease in the total distance travelled by households to the nearest clinic. After the initial decrease, all the scenarios have a general downward trend in the total distance travelled by households to the nearest clinic, with some fluctuation influenced by the number of new households in an area and the number of new clinics built. These results can be found in Table 6.3. For the relocation scenario, the decrease is much more gradual over the years. This gradual change can be attributed to the restricted budget that limits the number of clinics that can be opened. When comparing the total distance travelled by the households between the three scenarios, the values are as expected. The economic spike scenario has the shortest distances, followed by the trend scenario and lastly, the relocation scenario. This difference in the total distance travelled by households can mainly be attributed to the number of households in the lower-income categories in the three scenarios, where the relocation scenario has the most households and the economic spike scenario has the least households in the lower-income categories.

The accessibility improvement over the years for all three scenarios are depicted in Figure 6.3. By 2024, the overall accessibility will have improved significantly in all three scenarios. Tshwane will have the most households further than 5 km from the nearest clinic, however, this is also the municipality that had the worst accessibility to start with. The accessibility heat maps of the trend and relocation scenarios in 2024 look similar as both have more households on the peripheries of the municipalities than the economic spike scenario that has more households located closer to the city centres. Therefore, the economic spike scenario first focuses on locating clinics where the most households are and later cater for the less dense households further away. By 2030, the accessibility in Tshwane will also have improved, leaving only a few households in the three metros further than 5 km from the nearest clinics but closer than 10 km.

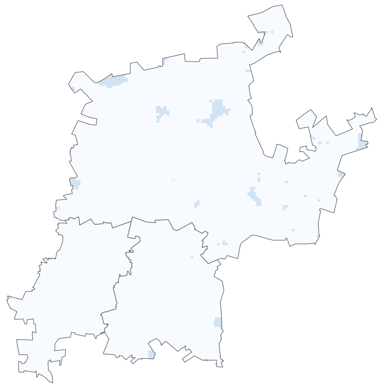
The distance from the distribution centre to the opened clinics in all three scenarios can be seen in Table 6.4. The total distance is a function of the number of clinics opened and the distance from the distribution centre. The total distance for the relocation scenario is the highest because this scenario has the most additional clinics opened and more of the clinics are on the peripheries of the municipalities than in the other scenarios due to the population growth and distribution. The economic spike scenario has the shortest total distance to the clinics as



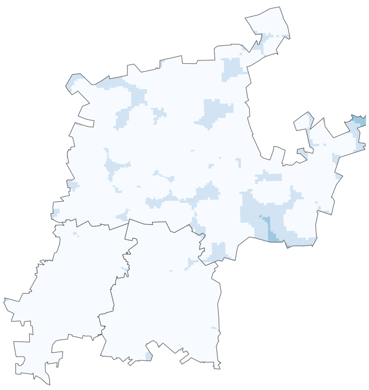
(a) 2018 (base line)



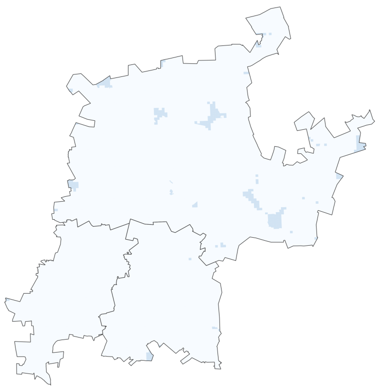
(b) 2024 (trend scenario)



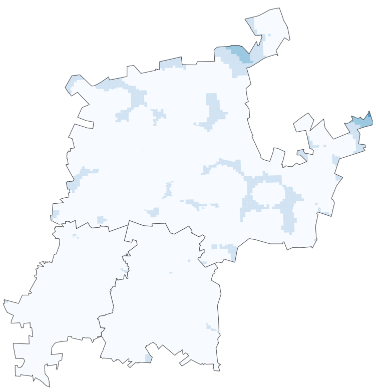
(c) 2030 (trend scenario)



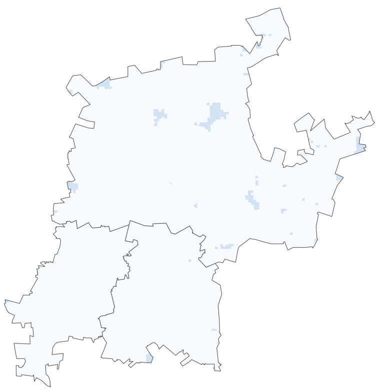
(d) 2024 (economic spike scenario)



(e) 2030 (economic spike scenario)



(f) 2024 (relocation scenario)



(g) 2030 (relocation scenario)

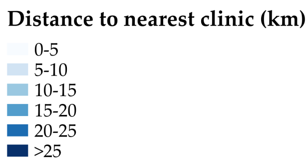


Figure 6.3: Accessibility improvement over the years for the three scenarios with budget constraints

Table 6.3: The total distance travelled by households to the nearest clinic per year for all three scenarios per year

Year	Trend scenario (km)	Economic spike scenario (km)	Relocation scenario (km)
2018	16 689	16 689	16 689
2019	10 110	10 747	15 564
2020	9 199	9 320	13 916
2021	8 731	9 007	13 375
2022	8 538	8 524	12 556
2023	8 649	8 472	12 340
2024	8 691	8 350	12 198
2025	8 600	8 199	12 216
2026	8 591	7 941	12 240
2027	8 574	8 034	12 293
2028	8 628	8 107	12 217
2029	8 677	8 130	12 210
2030	8 677	8 130	12 210

the clinics in this scenario are closer to the city centres, and fewer clinics are opened in this scenario than in the other two scenarios. All three scenarios follow a normal distribution for the distance from the distribution centre to all the opened clinics. These normal distributions confirm that the distribution is centrally located with regard to the clinics. The distribution centre is ideally located to service the clinics in the three metros. The mean travel distance between the distribution centre and the open clinics for all the scenarios is between 40 km and 43 km. These distances are very similar to the original case investigated as the location of the clinics have not changed much; it is mostly the year in which the clinics are opened that changed by introducing the budget constraint. The demand is still at the same nodes, and therefore clinics are still required in the same locations as the base case.

Table 6.4: Total distance from distribution centre to opened clinics for all three scenarios per year

Year	Trend scenario (km)	Economic spike scenario (km)	Relocation scenario (km)
2018	0	0	0
2019	24 481	20 095	15 500
2020	23 277	12 968	23 476
2021	25 225	27 376	21 760
2022	19 227	22 565	18 116
2023	16 323	19 319	20 013
2024	11 922	18 631	21 630
2025	19 612	17 423	29 025
2026	19 686	16 183	28 077
2027	24 748	15 610	14 296
2028	19 795	23 363	24 230
2029	22 936	26 373	21 564
2030	0	0	0

The robust solution aimed to come as close as possible to all three the individual solutions, as described above. The objectives of the individual scenarios were set as the goals in the goal programming model to find a robust configuration. The robust model placed facilities over the years to come as close to the individual scenarios as possible. The robust model seeks the best compromise between the three scenarios. The results of the robust solution are provided in Table 6.5. The robust model also adheres to the budget, forcing a more equal distribution of clinics to be opened over the years. The total cost for the robust model is around 8% higher than the scenario cost in any of the other scenarios, showing that robustness comes at a cost. The difference between the scenario values and the values given the robust configuration are within 10%. This 10% difference between the scenario values and the robust values is better than the 20% difference in the original case. The smaller difference is because the restricting budget ensures a more even placement of the clinics over the years and not at a surge placement in the first year. Therefore, even with additional constraints and changing parameters, the model can find a robust solution.

Table 6.5: Budget scenario robust configuration yearly solutions

Year	Total cost (R)	Trend scenario total distance travelled by households to the nearest clinic (km)	Economic spike scenario total distance travelled by households to the nearest clinic (km)	Relocation scenario total distance travelled by households to the nearest clinic (km)	Total distance from distribution centre to opened clinics (km)
2018	0	16 689 239	16 689 239	16 689 239	0
2019	7 823 499	10 019 787	10 129 638	15 542 152	42 867
2020	9 017 471	8 899 738	89 62 004	13 494 607	20 648
2021	7 999 785	8 572 304	8 553 000	12 727 732	23 265
2022	9 462 603	8 388 241	8 219 435	12 084 745	21 053
2023	7 957 458	8 564 024	8 279 381	12 067 717	25 034
2024	9 503 041	8 618 613	8 341 883	12 120 807	19 806
2025	6 775 668	8 452 000	8 223 803	11 872 826	26 207
2026	11 084 775	8 427 909	8 235 647	11 803 367	27 313
2027	11 472 742	8 214 731	8 036 109	11 505 891	22 793
2028	11 389 623	8 052 017	7 918 599	11 266 171	27 737
2029	14 547 214	8 010 768	7 886 244	11 162 252	27 589
2030	0	8 010 768	7 886 244	11 162 252	0

The model can find a robust configuration that is relatively close to the scenario solutions with a restrictive budget constraint. The model can be used to provide decision support for the health care department to identify feasible locations for clinics that will improve the overall accessibility of primary health care by finding the best locations and years to open clinics to cope with the growing demand. With a constrained budget, the model provides a strategic plan of when and where to open clinics in order to improve the overall accessibility of health care for the lower-income households. The available budget for new clinics can be used as input in the models to determine the best robust configuration of clinics to provide the desired level of accessibility.

6.3 Concluding remarks

Different accessibility constraints were tested and pareto frontiers were created with the associated variable values. The cost increased as the accessibility constraint was tightened. The total distance travelled by households decreased as the accessibility percentage increased. Lastly, the total distance from the distribution centre to the open clinics increased as the accessibility constraint was tightened as more clinics were opened. All these variables performed as expected and these pareto frontiers can be used to select an appropriate accessibility percentage. A budget analysis was done where a constraint was placed on the budget and the five-year accessibility constraint was increased to 12 years. The model successfully adapted to the restrictive budget and spread the opening of clinics across all the years. The robust model was within 10% of the scenario values, showing that the model can find a robust solution with an additional budget constraint. Concluding remarks and recommendations are made in Chapter 7.

Chapter 7

Conclusion

Health care, and especially access to health care, has always been a key metric for countries. In 2017, 9% of South Africa's GDP was spent on health care. Despite this being 5% higher than recommended by the World Health Organisation for a country of its socio-economic status, the country's health care status is poor in comparison to similar countries. In 1994, South Africa implemented a health care policy to make health care accessible to all South Africans. Accessible is defined as primary health care no more than 5 km from a place of residence. There is still a significant gap between the actual and desired accessibility for the lower-income communities and a need to improve access to public health care for all South Africans. Cost-effective and sustainable solutions are required to solve this problem. Therefore, an opportunity was identified to investigate the location of low-cost container clinics in lower-income communities.

There is a lack of dynamic location models that consider the changes in the problem environment over time, such as patient population and population migration. This project aimed to assist in closing this gap in the literature, using robust optimisation and goal programming to locate health care facilities in an uncertain environment using multiple scenarios.

This study investigated three development scenarios in the three metro municipalities (City of Tshwane, City of Johannesburg, and City of Ekurhuleni) in Gauteng, South Africa. The first scenario is a continuation of the current development and population growth trends. The second scenario sees a spike in economic development, leading to a reduction in lower-income households. The last scenario sees a rapid increase in lower-income households due to the relocation of households from rural areas to the three metros.

Associative forecasting methods were investigated to make more accurate health care demand forecasts based on expected changes in major cities due to urbanisation and economic growth. Household income, number of children, and the distance to the nearest clinics were the selected attributes for the demand forecasts. These factors were used to calculate the probability of a household member visiting a clinic when ill.

This demand and the location of the households were used as input into the facility location models. A good configuration of clinics was determined for each scenario using a Genetic Algorithm (GA). After that, the scenario values of each scenario were used as the goals in the goal programming model to determine a robust configuration that will work relatively well in all the scenarios given the uncertainty of the future development of the metros. A deviation of 25% was defined as an acceptable deviation from the individual scenario values since the scenarios investigated are so different.

For the individual scenarios, a good total distance to the nearest clinic for all the households has the same proportions as the total number of households. The total distance from the distribution centre to the open clinics are more or less the same in all the scenarios because the distribution centre is located centrally. The total costs for the scenarios are also very similar as most of the clinics are opened in the first year to serve the immediate need. There is a less than 20% configuration overlap between the different scenarios. Therefore, the solution for one scenario is not good or even close to acceptable for another scenario.

The robust model, utilising goal programming, was run to determine a robust solution that is as close as possible to the variable values of each scenario while minimising the standard deviation between the values. The difference between the scenario values and the robust solution is within a 25% range in all the scenarios. This overall variance is acceptable for a robust solution in an uncertain environment. The robust solution has an almost

40% configuration overlap with all the individual scenarios. These configuration overlaps are significantly higher than the just under 20% overlaps that are seen between the individual scenarios. Therefore, it was concluded that the robust configuration is a good compromise between the different scenario.

City planners, the Department of Health and other key role players, can use these models as quantitative decision support when deciding when and where to open new container clinics while they are still uncertain of which development strategy will realise in the next 10 to 15 years to achieve their accessibility goal. City planners, the Department of Health and other key role players, can look specifically at the total household travel distance and the total cost to determine the feasibility of the robust solution, while pharmaceutical or logistics companies can do their strategic planning using the total distribution distance.

An accessibility analysis was conducted to look at the impact of the accessibility target on model results. Three accessibility targets were selected to run the model. The model produced results as expected. The total cost increased as the accessibility increased. The household travel distance decreased as the accessibility increased as fewer households are further than 5 km from the nearest clinic. The benefits gained from the changed accessibility can be weighed up against the cost of reaching that accessibility target. These accessibility scenarios can be used for decision support when choosing an acceptable accessibility target.

Another analysis was conducted where the number of clinics that can be built per year was constrained by introducing a budget constraint. In this case, accessibility improvement is spread more evenly over the 12 years as the opening of clinics is also spread more evenly over the years. By incorporating a budget constraint, the model can assist decision makers in determining when to open which clinics while maximising the benefits to the community. Based on the results of the analyses, it was concluded that the model is useful for different scenarios with changes in parameters and that the model can be used for different scenarios.

7.1 Future work

The work in this report examines a simplified case of urban planning and robust facility location. Future work can include removing some of the assumptions made to simplify the model. Instead of using straight line distance between the households and clinics, actual road networks can be used. This will ensure greater accuracy in the distances travelled by households to the clinics and the distance from the distribution centre to the clinics.

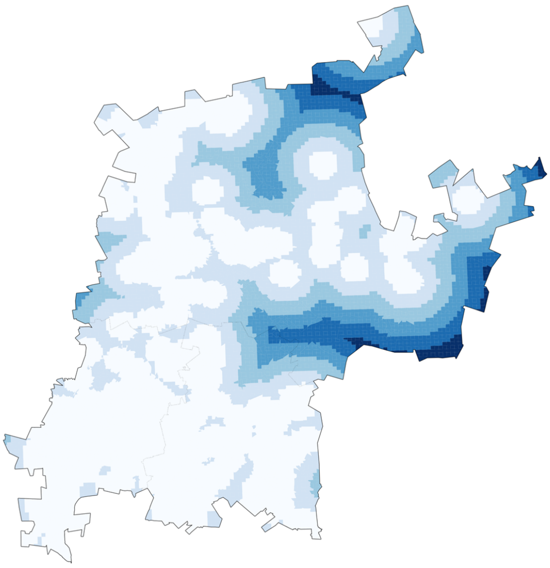
Another opportunity is to vary the building cost based on the location of the available land. The cost of land in prominent areas or close to city centres is higher than land in less dense and rural areas. By adding a land cost component to the building costs, a picture closer resembling reality can be created. This cost component will definitely impact the location of clinics when the cost is one of the trade-off points.

In future work, research on container clinics' operating costs can be done and the assumption of no operating costs can be removed. By adding operating costs, clinics will be built only when needed to ensure that the total cost is minimised. By including the operating costs in the model, it will be a closer representation of the real world. The model will no longer place most of the clinics in the first year as there will be a penalty in the form of operating costs if a clinic is opened before it is needed.

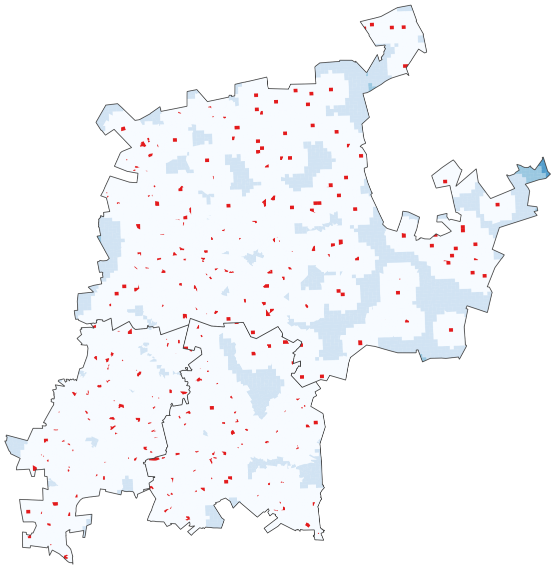
Further research that can be done includes incorporating different types of facilities such as hospitals and community health centres. These different facilities have different capacities and services that they can provide. Lastly, distribution centres and clinics can be located simultaneously to find the good locations or configurations for the scenarios and robust locations or configurations.

Appendices

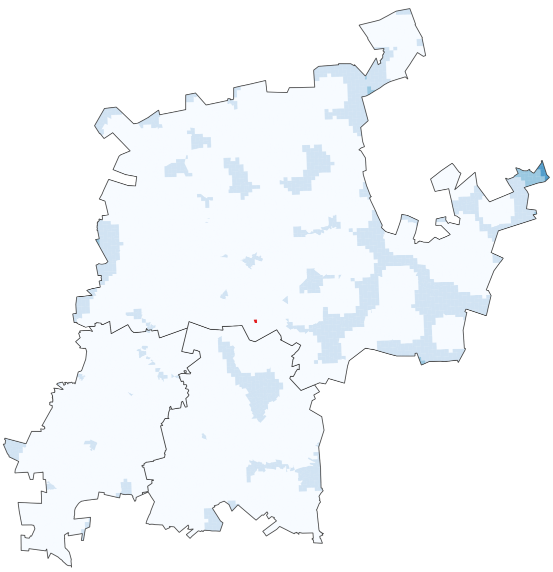
A Trend scenario facility locations



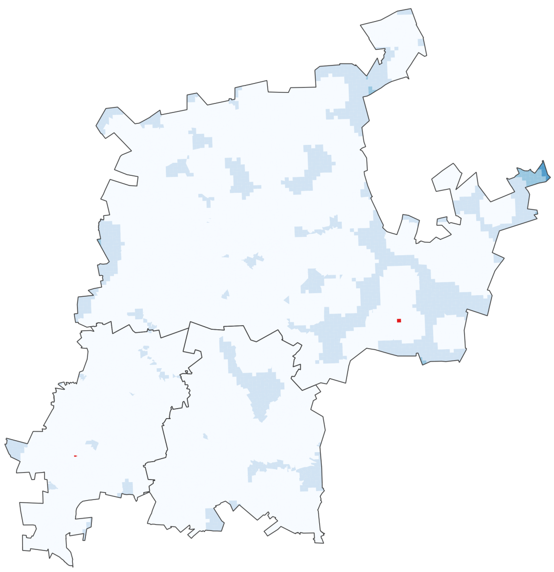
(a) 2018



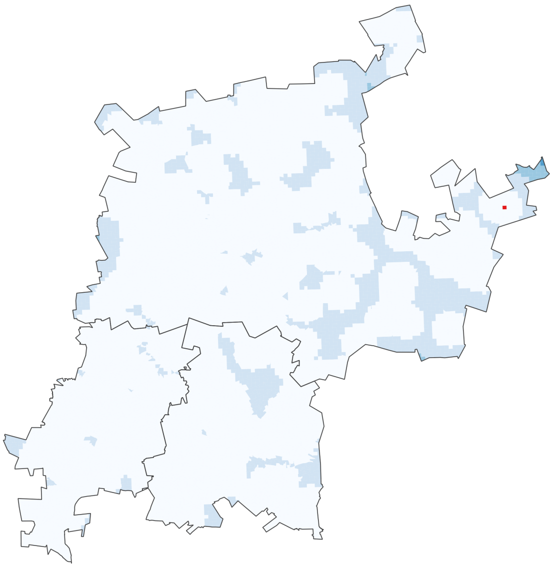
(b) 2019



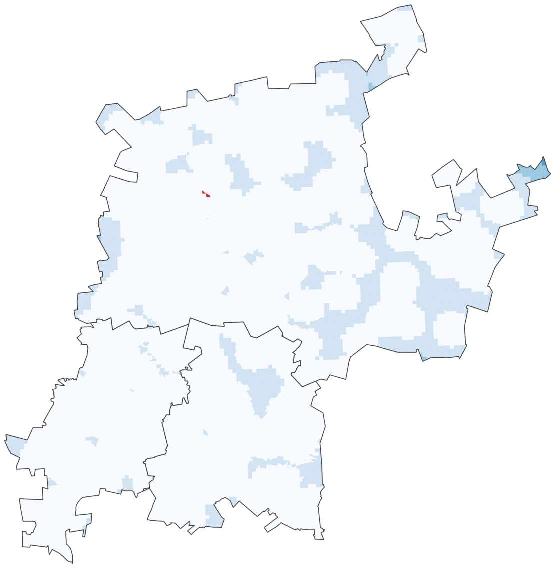
(c) 2020



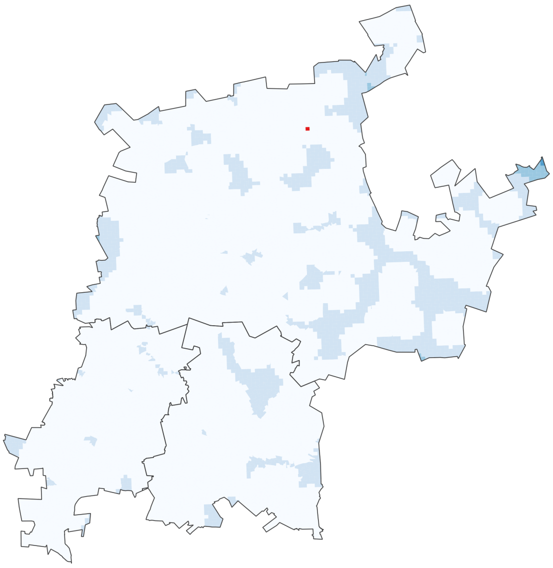
(d) 2022



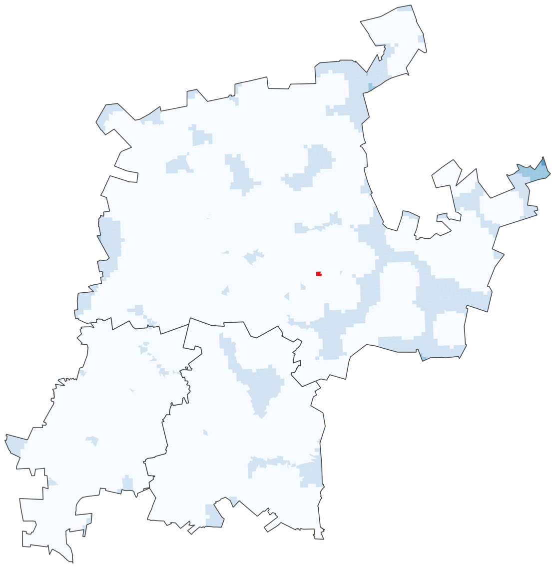
(e) 2024



(f) 2027



(g) 2028



(h) 2029

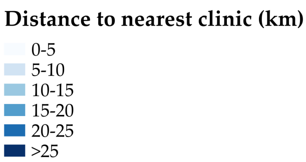
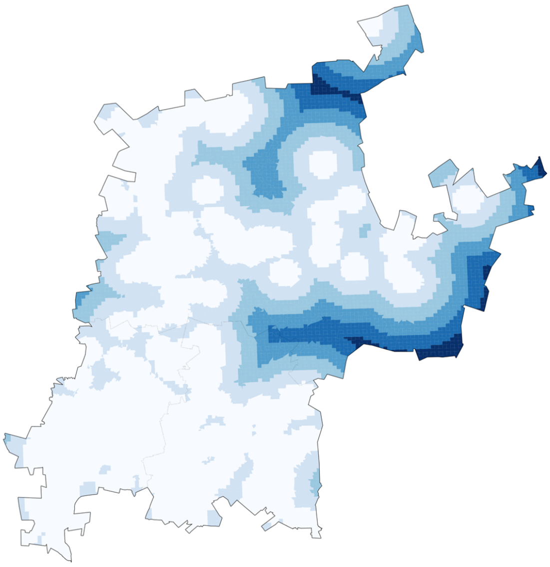
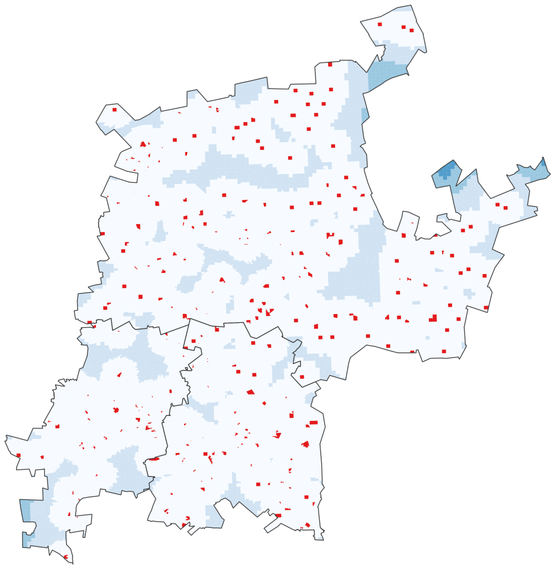


Figure A.1: Location of the clinics to opened for the trend scenario per year

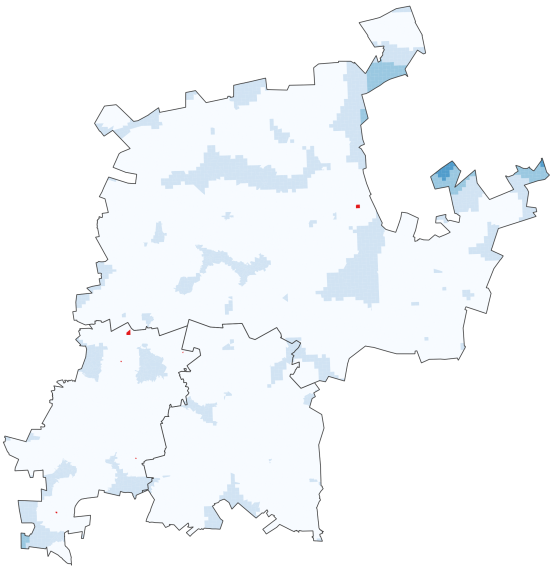
B Economic spike scenario facility locations



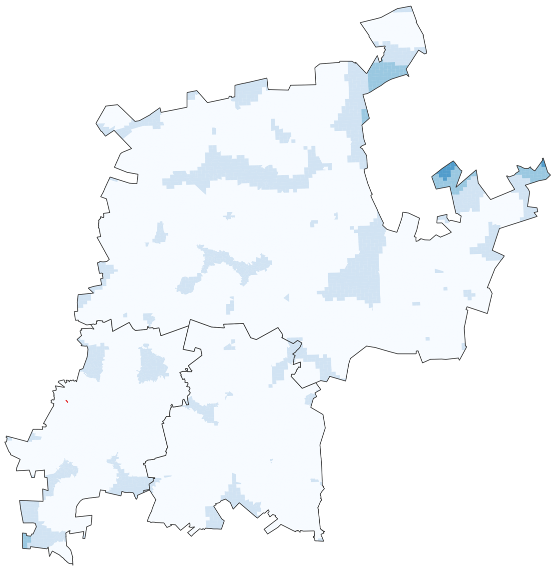
(a) 2018



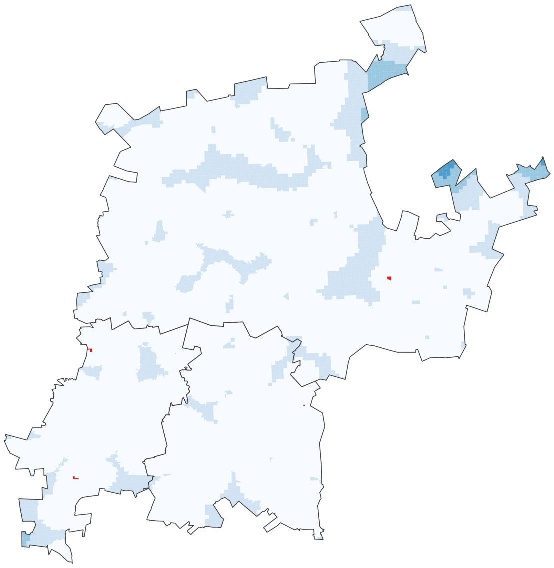
(b) 2019



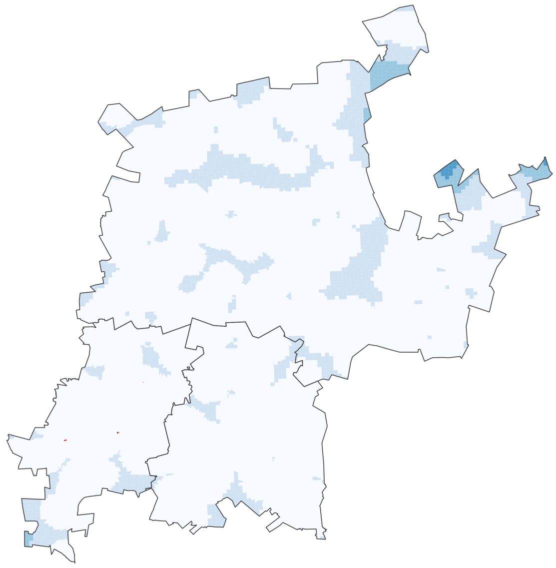
(c) 2020



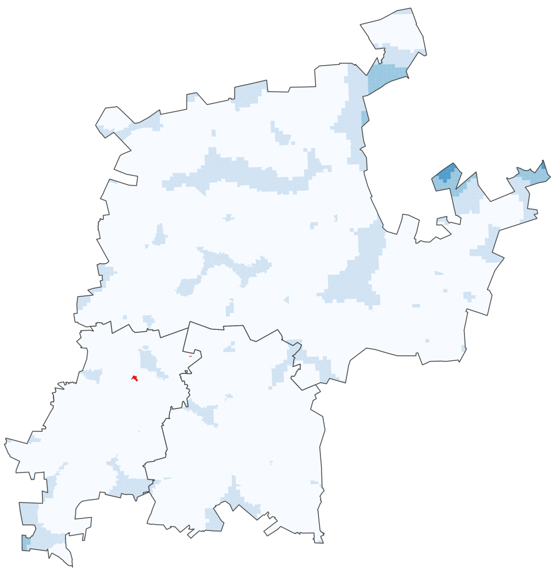
(d) 2022



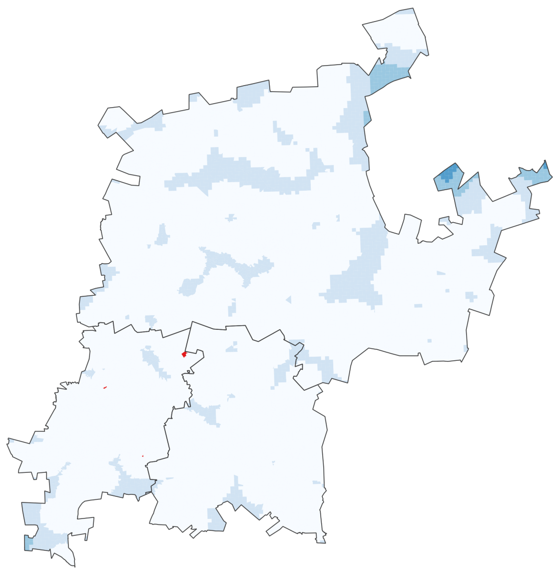
(e) 2023



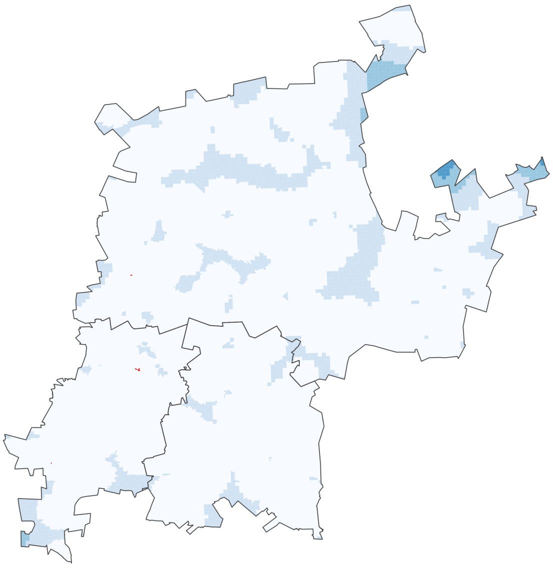
(f) 2025



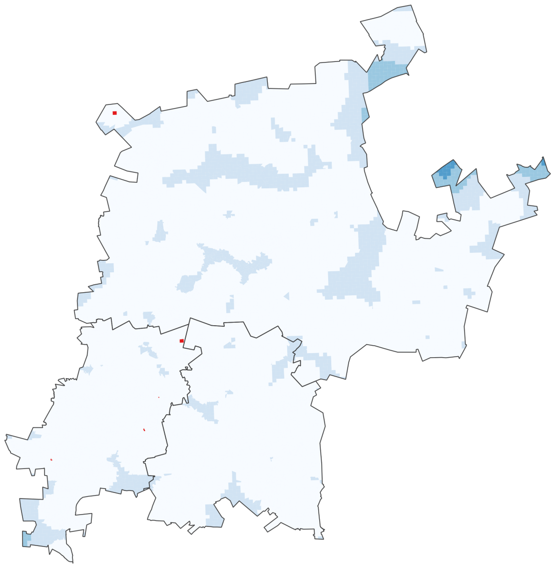
(g) 2026



(h) 2027



(i) 2028



(j) 2030

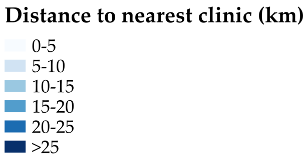
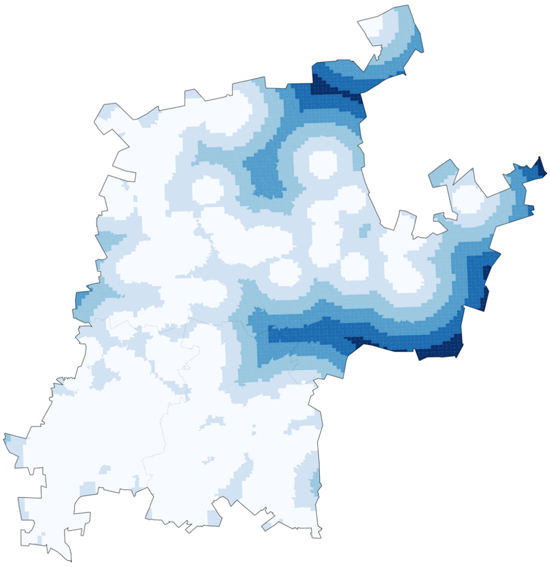
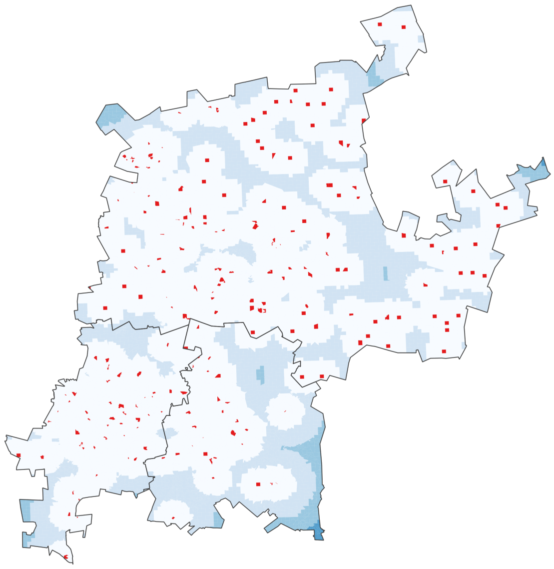


Figure B.1: Location of the clinics to be opened for the economic spike scenario per year

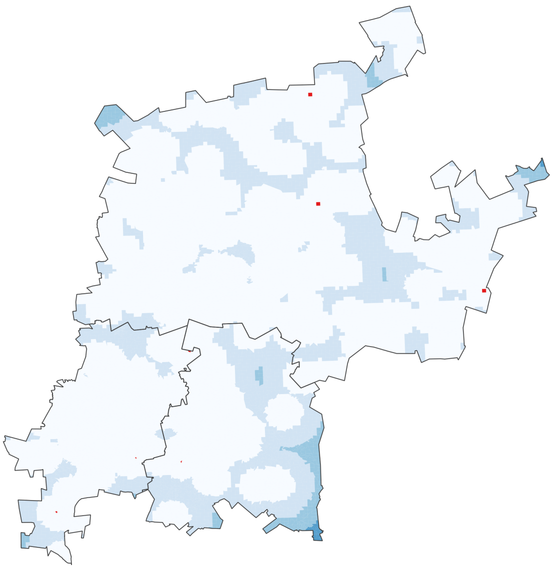
C Relocation scenario facility locations



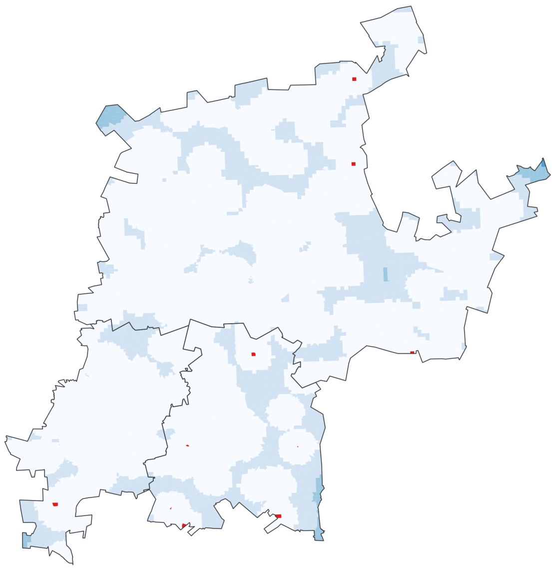
(a) 2018



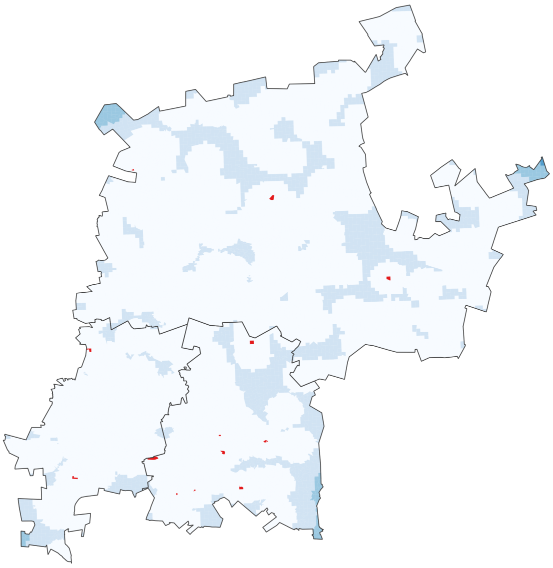
(b) 2019



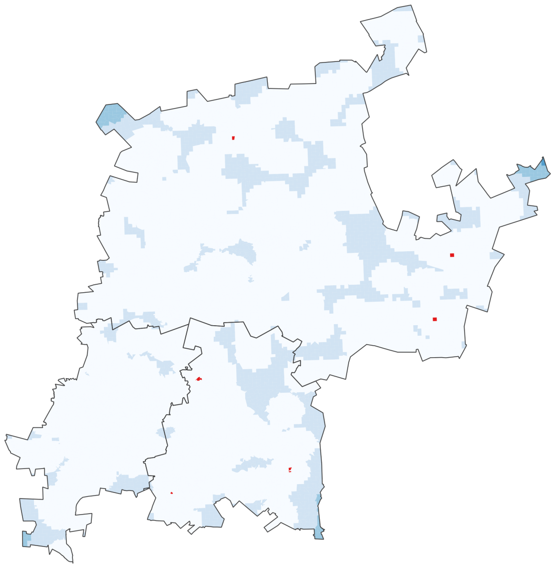
(c) 2020



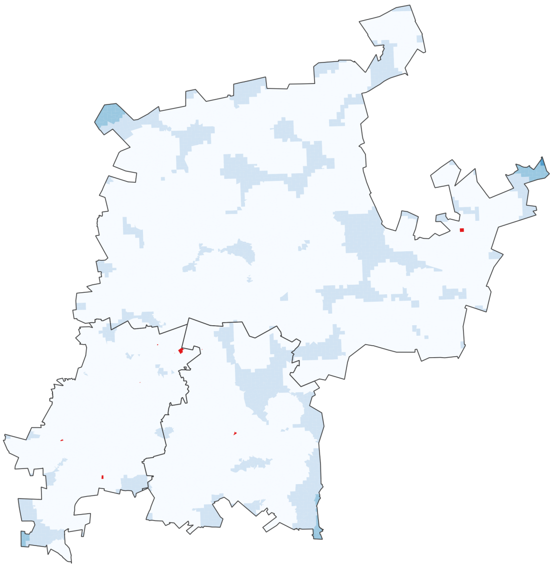
(d) 2021



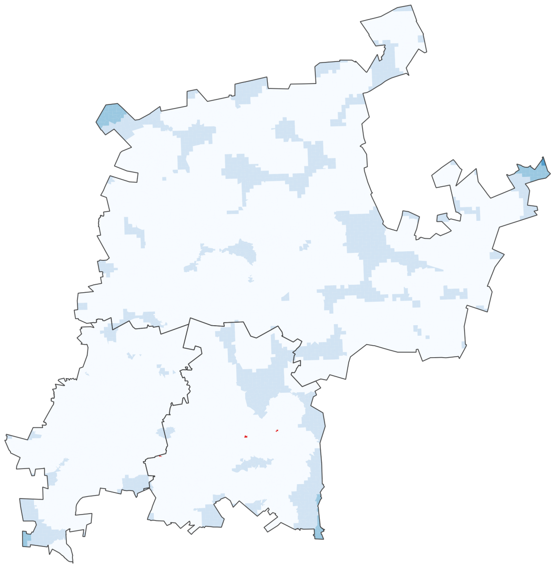
(e) 2023



(f) 2024



(g) 2025



(h) 2026

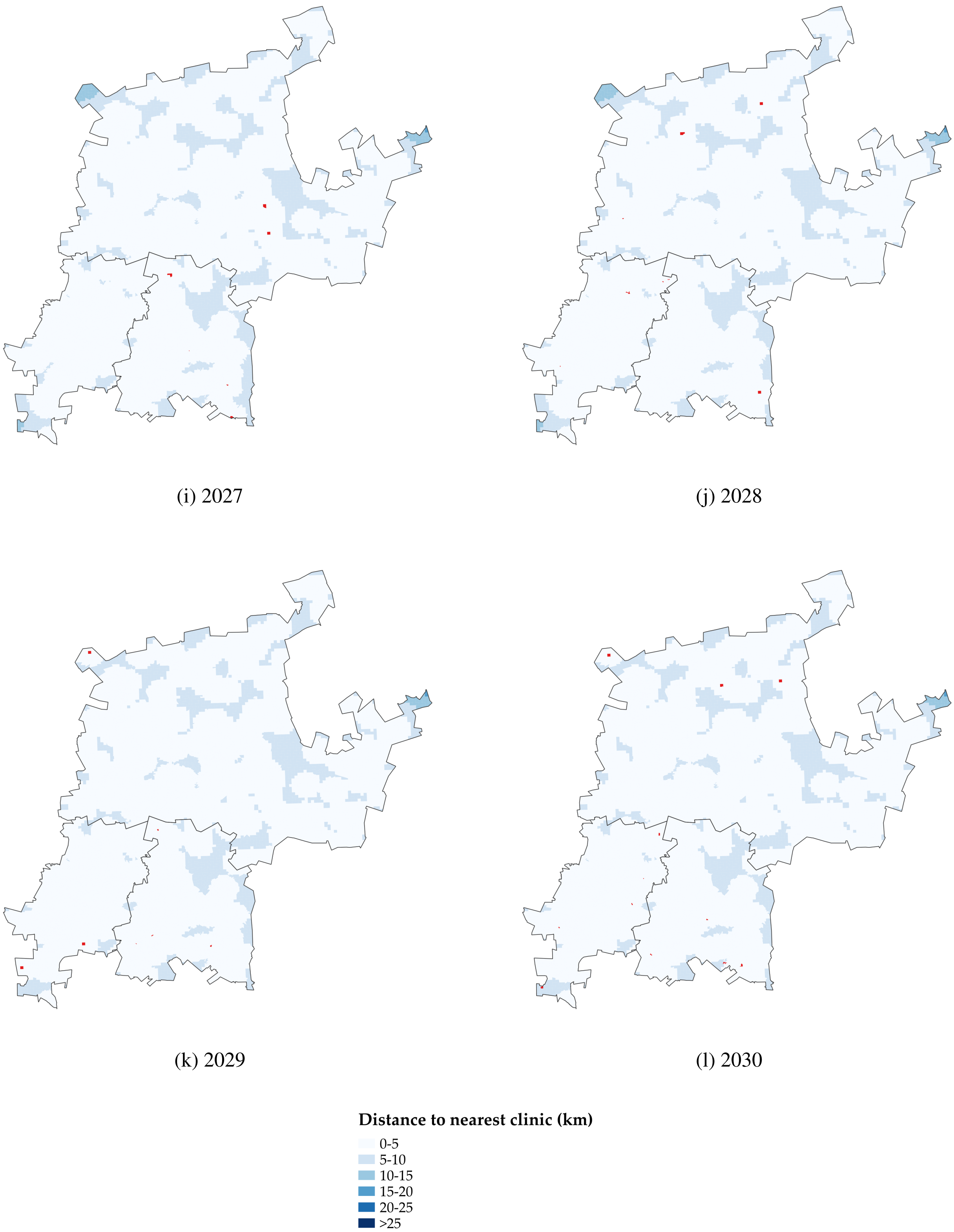
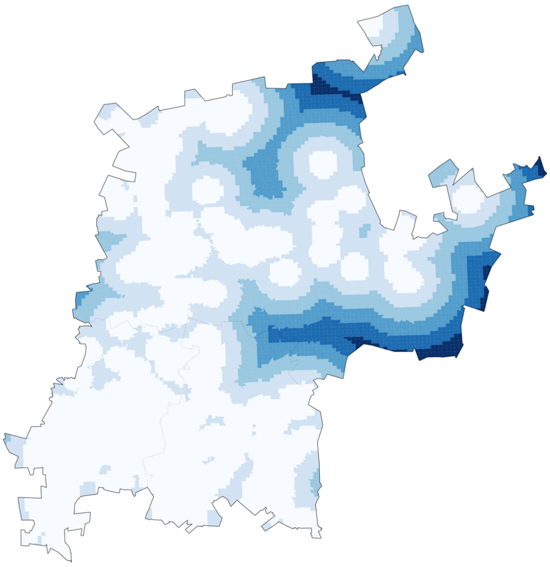
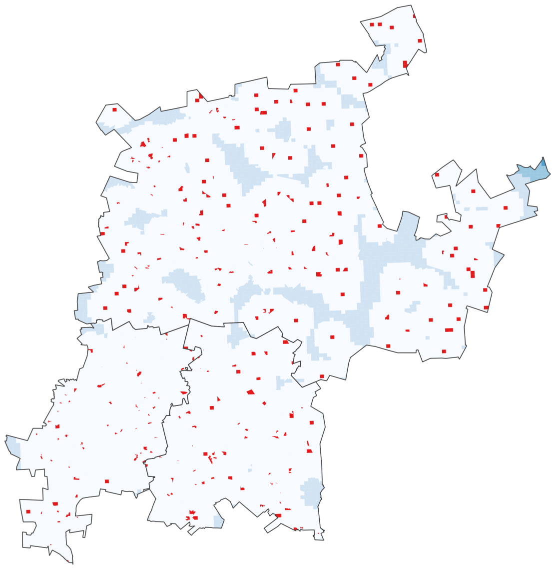


Figure C.1: Location of the clinics to be opened for the relocation scenario per year

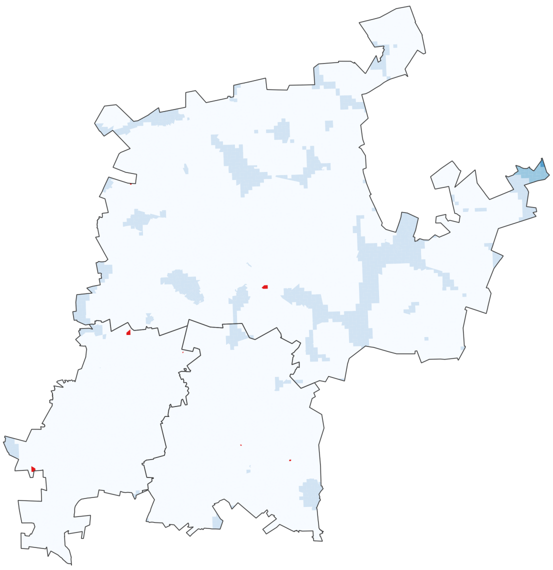
D Robust facility locations



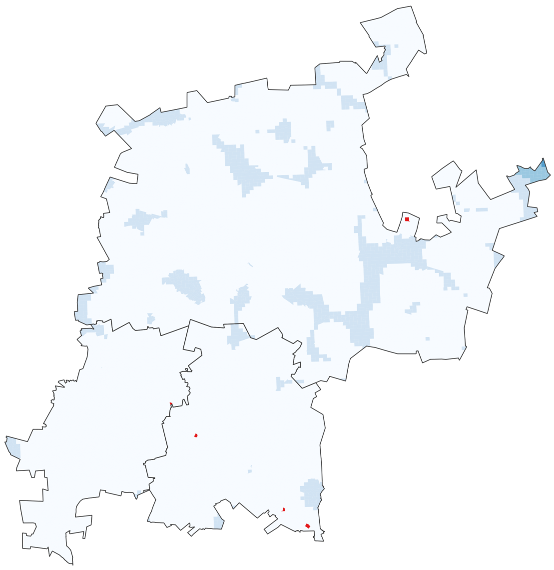
(a) 2018



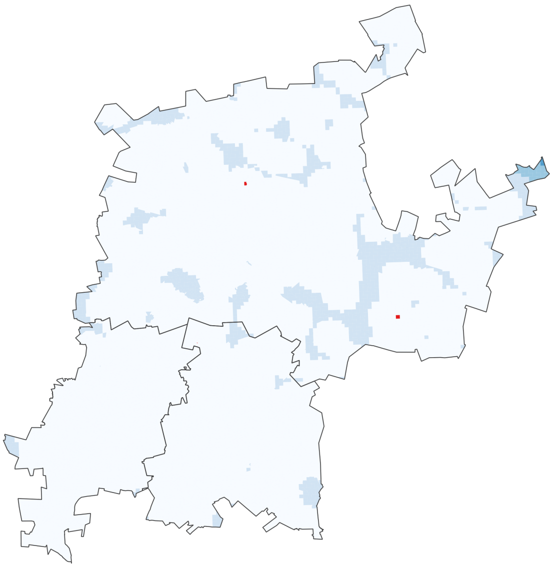
(b) 2019



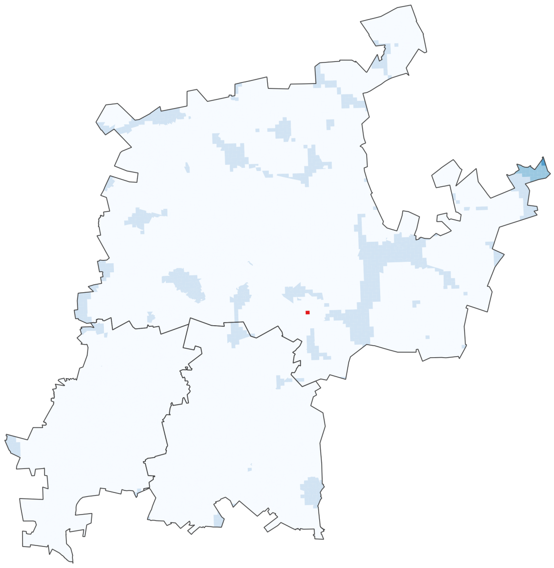
(c) 2020



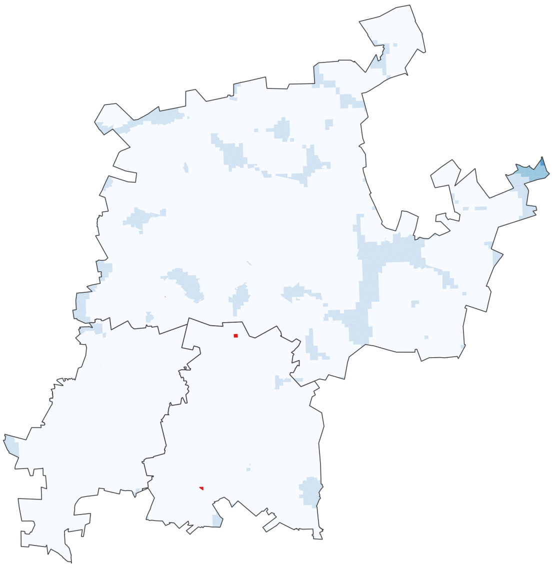
(d) 2021



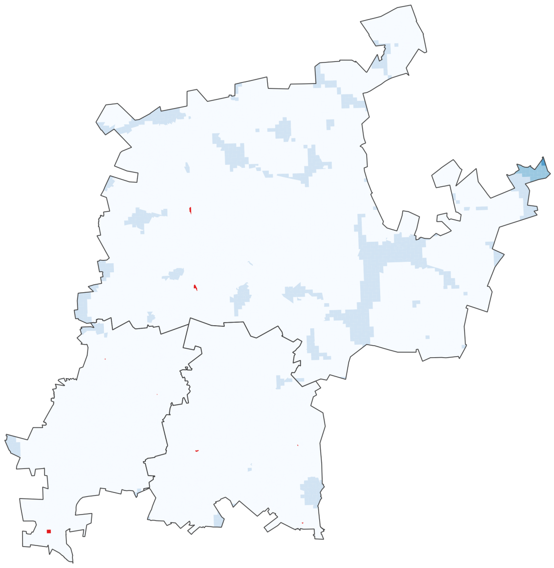
(e) 2022



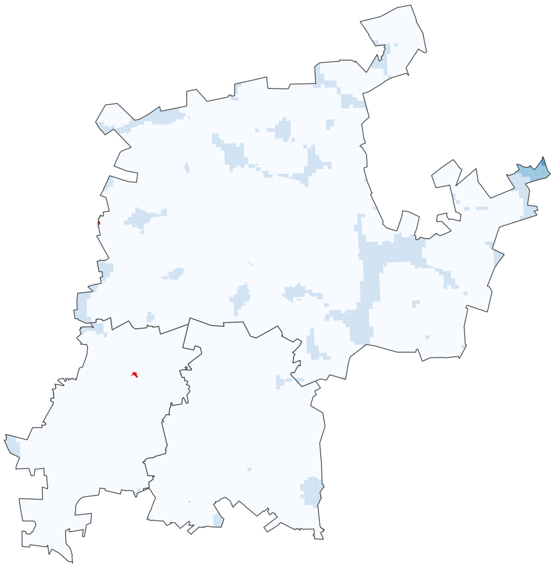
(f) 2023



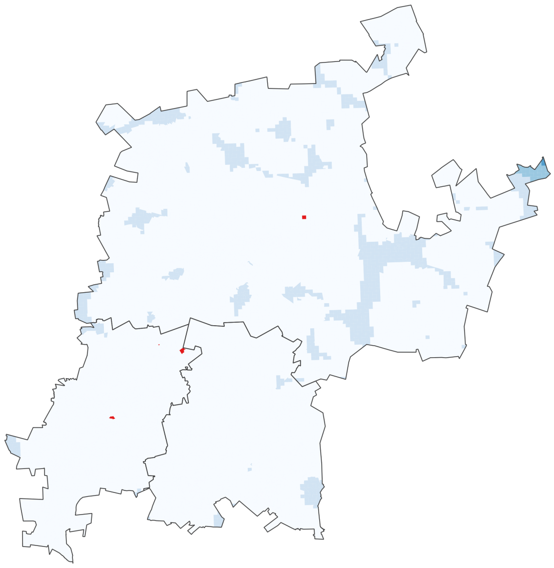
(g) 2024



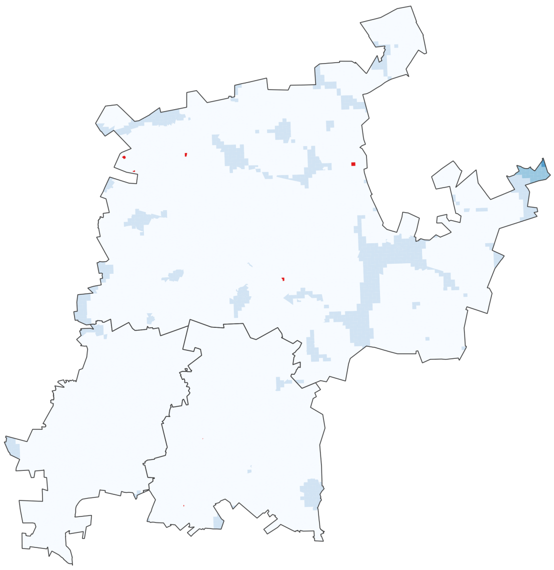
(h) 2025



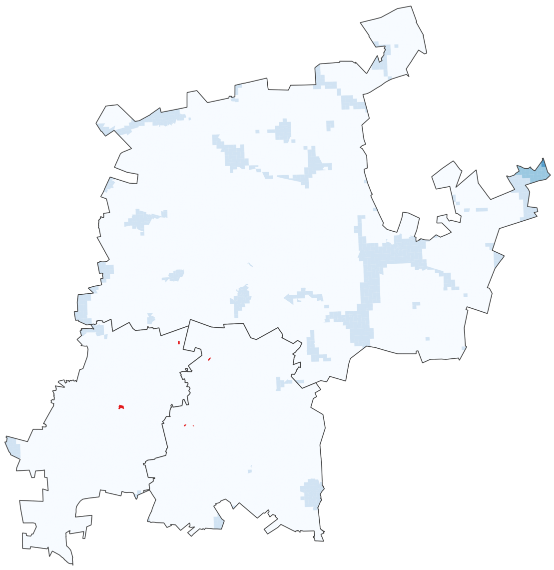
(i) 2026



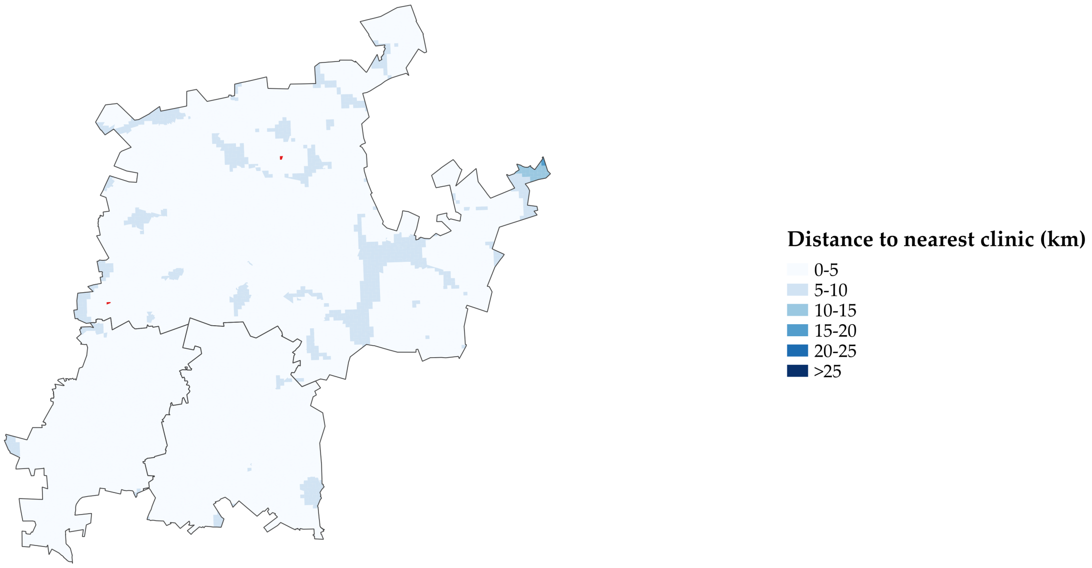
(j) 2027



(k) 2028



(l) 2029



(m) 2030

Figure D.1: Locations of the opened clinics of robust configuration per year

Bibliography

Abera Abaerei, A., Ncayiyana, J., and Levin, J. (2017). Health-care utilization and associated factors in Gauteng province, South Africa. *Global health action*, 10(1):1305765.

Africa Health (2019). *Industry Insights: South Africa Healthcare Market Overview*. Africa Health.

Ahmadi-Javid, A., Seyedi, P., and Syam, S. S. (2017). A survey of healthcare facility location. *Computers & Operations Research*, 79:223–263.

Albright, B. (2007). An introduction to simulated annealing. *College Mathematics Journal*, 38(1):37–42.

Arostegui, M. A., Kadipasaoglu, S. N., and Khumawala, B. M. (2006). An empirical comparison of tabu search, simulated annealing, and genetic algorithms for facilities location problems. *International Journal of Production Economics*, 103(2):742–754.

Aytug, H. and Saydam, C. (2002). Solving large-scale maximum expected covering location problems by genetic algorithms: A comparative study. *European Journal of Operational Research*, 141(3):480–494.

Baron, O., Milner, J., and Naseraldin, H. (2011). Facility location: A robust optimization approach. *Production and Operations Management*, 20(5):772–785.

Barthelemy, M. (2011). Spatial networks. *Physics Reports*, 499(1-3):1–101.

Beheshtifar, S. and Alimohammadi, A. (2014). A multiobjective optimization approach for location-allocation of clinics. *International Transactions in Operational Research*, 22.

Big Box Containers (2020). Shipping Container Clinics: A Clever Solution. https://www.bigboxcontainers.co.za/blog/shipping-container-clinics-a-clever-solution. Accessed: 2019-08-05.

Booysen, F. (2003). Urban-rural inequalities in health care delivery in South Africa. *Development Southern Africa*, 20(5):659–673.

Buor, D. (2003). Analysing the primacy of distance in the utilization of health services in the Ahafo-Ano South district, Ghana. *The International journal of health planning and management*, 18(4):293–311.

Bureau for economic research (2019). Building cost report - quarterly analysis of building costs. Technical report.

Cebecauer, M. and Buzna, L. (2017). A versatile adaptive aggregation framework for spatially large discrete location-allocation problems. *Computers & Industrial Engineering*, 111:364–380.

Chang, C. (2015). Multi-choice goal programming model for the optimal location of renewable energy facilities. *Renewable and Sustainable Energy Reviews*, 41(1):379–389.

Chiyoshi, F. and Galvao, R. D. (2000). A statistical analysis of simulated annealing applied to the p-median problem. *Annals of Operations Research*, 96(1-4):61–74.

Church, R. and ReVelle, C. (1974). The maximal covering location problem. *Papers of the Regional Science Association*, 32(1):101–118.

Clarke, G. and Wright, J. W. (1964). Scheduling of vehicles from a central depot to a number of delivery points. *Operations Research*, 12(4):568–581.

Cooke, G. S., Tanser, F. C., Bärnighausen, T. W., and Newell, M. (2010). Population uptake of antiretroviral treatment through primary care in rural south africa. *BMC public health*, 10(1):585.

Das, S. K., Roy, S. K., and Weber, G. W. (2020). An exact and a heuristic approach for the transportation-p-facility location problem. *Computational Management Science*, 17(3):389–407.

Daskin, M. S. (2008). What you should know about location modeling. *Naval Research Logistics*, 55(4):283–294.

Daskin, M. S., Hesse, S. M., and Revelle, C. S. (1997). α-reliable p-minimax regret: A new model for strategic facility location modeling. *Location Science*, 5(4):227–246.

Department of Health (2018). Primary health care facilities and services. https://www.healthestablishments.org.za/Home/Facility. Accessed: 2019-08-14.

Du, B. and Zhou, H. (2018). A robust optimization approach to the multiple allocation p-center facility location problem. *Symmetry*, 10(11):588.

Ekurhuleni Metropolitan Municipality (2015). Integrated Development Plan 2014/15. https://www.ekurhuleni.gov.za/council/public-engagements/idp/482-idp-budget-and-sdbip-201314-201516. Accessed: 2020-05-11.

Farahani, R. Z., SteadieSeifi, M., and Asgari, N. (2010). Multiple criteria facility location problems: A survey. *Applied Mathematical Modelling*, 34(7):1689–1709.

Fortin, F.-A., De Rainville, F.-M., Gardner, M.-A., Parizeau, M., and Gagné, C. (2012). DEAP: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, 13:2171–2175.

Gendreau, M. (2008). Metaheuristics for the vehicle routing problem and its extensions : a categorized bibliography. *Vehicle routing problem : latest advances and new challenges.*

Ghodratnama, A., Tavakkoli-Moghaddam, R., and Azaron, A. (2015). Robust and fuzzy goal programming optimization approaches for a novel multi-objective hub location-allocation problem: A supply chain overview. *Applied Soft Computing*, 37:255–276.

Güneş, E. D., Yaman, H., Çekyay, B., and Verter, V. (2014). Matching patient and physician preferences in designing a primary care facility network. *Journal of the Operational Research Society*, 65(4):483–496.

Graber-Naidich, A., Carter, M. W., and Verter, V. (2015). Primary care network development: the regulator's perspective. *Journal of the Operational Research Society*, 66(9):1519–1532.

Graham, G., Mehmood, R., and Coles, E. (2015). Exploring future cityscapes through urban logistics prototyping: a technical viewpoint. *Supply Chain Management: An International Journal*, 20(3):341–352.

Gumte, K. M., Pantula, P. D., Miriyala, S. S., and Mitra, K. (2021). Data driven robust optimization for handling uncertainty in supply chain planning models. *Chemical Engineering Science*, page 116889.

Hart, W. E., Laird, C., Watson, J.-P., Woodruff, D. L., Hackebeil, G. A., Nicholson, B. L., and Siirola, J. D. (2017). Pyomo – Optimization Modeling in Python. *Springer*.

Havenga, J., Simpson, Z., King, D., de Bod, A., and Braun, M. (2016). Logistics Barometer South Africa 2016. *Logistics Baometer*, pages 1–13.

Hill, C. F., Powers, B. W., Jain, S. H., Bennet, J., Vavasis, A., and Oriol, N. E. (2014). Mobile health clinics in the era of reform. *American Journal of Managed Care*, 20(3).

Hochbaum, D. S. and Pathria, A. (1998). Locating centers in a dynamically changing network, and related problems. *Location Science*, 6(1-4):243–256.

Hotchkiss, D. R. (1998). The tradeoff between price and quality of services in the Philippines. *Social science & medicine*, 46(2):227–242.

Irawan, C. A., Imran, A., and Luis, M. (2020). Solving the bi-objective capacitated p-median problem with multilevel capacities using compromise programming and vns. *International Transactions in Operational Research*, 27(1):361–380.

IUSS (2014). Building Clinic Facilities with Innovative Building Technologies.

Jain, C. L. (2005). Benchmarking forecasting practices in corporate america. *The Journal of Business Forecasting*, 24(4).

Jones, S. S., Thomas, A., Evans, R. S., Welch, S. J., Haug, P. J., and Snow, G. L. (2008). Forecasting daily patient volumes in the emergency department. *Academic Emergency Medicine*, 15(2):159–170.

Katoch, S., Chauhan, S. S., and Kumar, V. (2020). A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, pages 1–36.

Kchaou Boujelben, M. and Boulaksil, Y. (2018). Modeling international facility location under uncertainty: A review, analysis, and insights. *IISE Transactions*, 50(6):535–551.

Kim, D.-G. and Kim, Y.-D. (2013). A lagrangian heuristic algorithm for a public healthcare facility location problem. *Annals of Operations Research*, 206(1):221–240.

Klein, M. G., Verter, V., and Moses, B. G. (2020). Designing a rural network of dialysis facilities. *European Journal of Operational Research*, 282(3):1088–1100.

Kuehn, A. A. and Hamburger, M. J. (1963). A heuristic program for locating warehouses. *Management science*, 9(4):643–666.

Levanova, T. and Gnusarev, A. (2018). Simulated annealing for competitive p-median facility location problem. In *Journal of Physics: Conference Series*, volume 1050, page 012044. IOP Publishing.

Macrotrends (2020). South Africa Population Growth Rate. https://www.macrotrends.net/countries/ZAF/south-africa/population-growth-rate. Accessed: 2019-06-15.

Marianov, V. and Taborga, P. (2001). Optimal location of public health centres which provide free and paid services. *Journal of the Operational Research Society*, 52(4):391–400.

Masiye, F. and Kaonga, O. (2016). Determinants of Healthcare Utilisation and Out-of-Pocket Payments in the Context of Free Public Primary Healthcare in Zambia. *International journal of health policy and management*, 5(28005549):693–703.

McLaren, Z. M., Ardington, C., and Leibbrandt, M. (2014). Distance decay and persistent health care disparities in South Africa. *BMC Health Services Research*, 14(1):541.

Meskarian, R., Penn, M. L., Williams, S., and Monks, T. (2017). A facility location model for analysis of current and future demand for sexual health services. *PloS one*, 12(8).

Mestre, A. M., Oliveira, M. D., and Barbosa-Póvoa, A. P. (2015). Location-allocation approaches for hospital network planning under uncertainty. *European Journal of Operational Research*, 240(3):791–806.

Miç, P., Koyuncu, M., and Hallak, J. (2019). Primary health care center (PHCC) location-allocation with multi-objective modelling: a case study in Idleb, Syria. *International journal of environmental research and public health*, 16(5):811.

Mitchell, S., Consulting, S. M., and Dunning, I. (2011). PuLP: A Linear Programming Toolkit for Python.

Mitropoulos, P., Mitropoulos, I., and Giannikos, I. (2013). Combining dea with location analysis for the effective consolidation of services in the health sector. *Computers & Operations Research*, 40(9):2241–2250.

Mwabu, G., Ainsworth, M., and Nyamete, A. (1993). Quality of medical care and choice of medical treatment in kenya: an empirical analysis. *Journal of Human Resources*, pages 838–862.

Nahu, A. (2006). Determinants of demand for health care services and their implication on Health care financing: the case of Bure town. *Ethiopian Journal of Economics*, 11.

Nemet, G. F. and Bailey, A. J. (2000). Distance and health care utilization among the rural elderly. *Social Science & Medicine*, 50(9):1197–1208.

Nteta, T. P., Mokgatle-Nthabu, M., and Oguntibeju, O. O. (2010). Utilization of the primary health care services in the Tshwane Region of Gauteng Province, South Africa. *PloS one*, 5(21085475):e13909–e13909.

O'donnell, O. (2007). Access to health care in developing countries: breaking down demand side barriers. *Cadernos de saude publica*, 23(12):2820–2834.

Peng, Q. and Afshari, H. (2014). Challenges and solutions for location of healthcare facilities. *Industrial Engineering & Management*, 03.

Rahman, S. and Smith, D. (2000). Use of location-allocation models in health service development planning in developing nations. *European Journal of Operational Research*, 123(3):437–452.

Rajagopalan, H. K., Vergara, F. E., Saydam, C., and Xiao, J. (2007). Developing effective meta-heuristics for a probabilistic location model via experimental design. *European journal of operational research*, 177(1):83–101.

Reeves, C. (2003). Genetic algorithms. In *Handbook of metaheuristics*, pages 55–82. Springer.

ReVelle, C., Toregas, C., and Falkson, L. (1976). Applications of the location set-covering problem. *Geographical analysis*, 8(1):65–76.

Rosenhead, J., Elton, M., and Gupta, S. K. (1972). Robustness and optimality as criteria for strategic decisions. *Journal of the Operational Research Society*, 23(4):413–431.

Sarma, S. (2009). Demand for outpatient healthcare: empirical findings from rural India. *Applied health economics and health policy*, 7(4):265–77.

Shariff, S. R., Moin, N. H., and Omar, M. (2012). Location allocation modeling for healthcare facility planning in Malaysia. *Computers & Industrial Engineering*, 62(4):1000–1010.

Shishebori, D. and Yousefi Babadi, A. (2015). Robust and reliable medical services network design under uncertain environment and system disruptions. *Transportation Research Part E: Logistics and Transportation Review*, 77:268–288.

Snyder, L. V. (2006). Facility location under uncertainty: a review. *IIE transactions*, 38(7):547–564.

Soebiyanto, R. P., Adimi, F., and Kiang, R. K. (2010). Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. *PloS one*, 5(3):e9450.

South Africa (1996). Constitution of the Republic of South Africa. https://www.refworld.org/docid/3ae6b5de4.html. Accessed: 2020-09-14.

Stats SA (2011). South African census 2011. Technical report, Stats SA.

Stats SA (2015). Measuring household expenditure on public transport. http://www.statssa.gov.za/?p=5943. Accessed: 2020-08-14.

Stats SA (2018). General household survey. http://www.statssa.gov.za/publications/P0318. Accessed: 2020-06-14.

Stats SA (2020). Mid-year population estimates 2020.

Sun, M. (2012). A Tabu Search Heuristic Procedure for the Capacitated Facility Location Problem. *Journal of Heuristics*, 18.

Syam, S. S. and Côté, M. J. (2010). A location-allocation model for service providers with application to not-for-profit health care organizations. *Omega*, 38(3-4):157–166.

Taiwo, O. J. (2020). Maximal covering location problem (mclp) for the identification of potential optimal covid-19 testing facility sites in nigeria. *African Geographical Review*, pages 1–17.

Talbi, E. (2009). *Metaheuristics : from design to implementation*. John Wiley & Sons,, Hoboken, N.J. :.

Tank, P. (2017). What is urban growth, 2012. https://planningtank.com/urbanisation/what-is-urban-growth. Accessed: 2019-04-12.

Tanser, F., Gijsbertsen, B., and Herbst, K. (2006). Modelling and understanding primary health care accessibility and utilization in rural south africa: an exploration using a geographical information system. *Social science & medicine*, 63(3):691–705.

The World Bank Group (2015). Master Planning. https://urban-regeneration.worldbank.org/node/51. Accessed: 2019-11-14.

Verter, V. and Lapierre, S. D. (2002). Location of preventive health care facilities. *Annals of Operations Research*, 110(1-4):123–132.

Wabiri, N., Chersich, M., Shisana, O., Blaauw, D., Rees, H., and Dwane, N. (2016). Growing inequities in maternal health in South Africa: a comparison of serial national household surveys. *BMC pregnancy and childbirth*, 16(1):256.

Wellay, T., Gebreslassie, M., Mesele, M., Gebretinsae, H., Ayele, B., Tewelde, A., and Zewedie, Y. (2018). Demand for health care service and associated factors among patients in the community of Tsegedie District, Northern Ethiopia. *BMC health services research*, 18(1):697.

Wichapa, N. and Khokhajaikiat, P. (2017). Solving multi-objective facility location problem using the fuzzy analytical hierarchy process and goal programming: a case study on infectious waste disposal centers. *Operations Research Perspectives*, 4:39–48.

Xu, T., Ren, Y., Guo, L., Wang, X., Liang, L., and Wu, Y. (2021). Multi-objective robust optimization of active distribution networks considering uncertainties of photovoltaic. *International Journal of Electrical Power & Energy Systems*, 133:107197.

Yang, H., Morton, D. P., Bandi, C., and Dvijotham, K. (2021). Robust optimization for electricity generation. *INFORMS Journal on Computing*, 33(1):336–351.

Yantzi, N., Rosenberg, M. W., Burke, S. O., and Harrison, M. B. (2001). The impacts of distance to hospital on families with a child with a chronic condition. *Social Science & Medicine*, 52(12):1777–1791.

Young, M. (2016). Private vs. public healthcare in south africa.