HARM DE VRIES

Evidence-Based Optimization in Humanitarian Logistics

Evidence-Based Optimization in Humanitarian Logistics

Evidence-Based Optimization in Humanitarian Logistics

Evidence-based optimalisatie binnen de humanitaire logistiek

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
Rector Magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board.

The public defense shall be held on
Friday 10 November at 9:30 hrs

by

Harm de Vries
born in IJsselmuiden.

Erasmus University Rotterdam

Doctoral Committee

Doctoral dissertation supervisors:

Prof.dr. A.P.M. Wagelmans

Prof.dr. J.J. van de Klundert

Other members:

Prof.dr. M.L. Brandeau

Prof.dr. D. Huisman

Prof.dr. L.N. Van Wassenhove

Erasmus Research Institute of Management - ERIM
The joint research institute of the Rotterdam School of Management (RSM)
and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam
Internet: http://www.erim.eur.nl

ERIM Electronic Series Portal: http://repub.eur.nl/

ERIM PhD Series in Research in Management, 435
ERIM reference number: EPS-2017-435-LIS
ISBN 9789058924971

Design: PanArt, www.panart.nl

This publication (cover and interior) is printed by Tuijtel on recycled paper, BalanceSilk®.
The ink used is produced from renewable resources and alcohol free fountain solution. Certifications for the paper and the printing production process: Recycle, EU Ecolabel, FSC®, ISO14001.
More info: www.tuijtel.com

Acknowledgments

Writing a thesis and obtaining a PhD degree was not something I ever imagined myself doing in my life. Yet, here I am, writing the preliminary of a finished PhD thesis. Apparently I somehow ended up in the scientific world. And apparently it was sufficiently rewarding to keep going for four years.

Looking back, I can indeed only conclude that I had some of the most wonderful years of my life. These years were not just a matter of sitting behind a desk, reading and writing articles. They were filled with enriching experiences. I learned all kinds of new methodologies, analyzed messy and pressing real life problems, cracked my brains to mathematically analyze such problems, traveled all over the world to discuss findings and visit the field, taught and supervised students, developed a network of contacts in the humanitarian and academic world, met many wonderful people, and had a lot of fun. There are many people that, in one way or another, contributed to this great experience, and I would like to take this opportunity to thank them.

Albert and Joris, you offered me the chance to prove myself as an academic and to do research on the topic that has my heart. I realize that this is quite a luxury. You have also been true role models for me. I really admire your sharp minds, your ability to master all responsibilities you have, and the constructive and empathic atmosphere around you. Thanks a lot for everything.

I would also like to thank Luk. You offered me the possibility to visit INSEAD as a PhD student and provided me the fantastic chance to stay there as a postdoctoral researcher. You have showed confidence in me and gave me responsibilities that I

greatly enjoy. It is inspiring to work with someone having such dedication to do relevant and impactful research. You have been, and still are, a great mentor for me.

Next, I am thankful to my other co-authors – Dennis, José, Sake, Epco, Crispin, and Pascal – as well as the people from the organizations I have been working with. It has been a real pleasure to work with you. Dennis and José, your enthusiasm about our research project and the dedication you put into it made it very nice to work with you. Sake, Epco, Crispin, and Pascal, combining our different fields of expertise and methodological backgrounds was a very exciting and enriching experience.

Furthermore, I also would like to thank my paranymphs and office mates Kristiaan and Kevin for the wonderful time I had at H7. You are not only among the smartest people I know, you are also among the most sympathetic and authentic ones. You always knew how to create an exciting scientific atmosphere around you. And you always knew when it was time to move aside our work to have a good discussion or to simply have some fun.

I am also truly grateful to my other EUR colleagues. You are a unique group of people. One in which people really care about each other, one in which you can have a lot of fun, and at the same time one with a productive and excellent scientific atmosphere. I have great memories to the numerous coffee breaks, office sports events, discussions, game nights, office cooking events, and other activities.

Furthermore, I would like to thank my family and friends for being such wonderful people. You were always there with your support and provided the reason to stop working every now and then. I most especially would like to thank my parents and parents in law. Pa and Ma, thank you for the warm home you always offered. You have taught me to be ambitious and to finish what you have started. It has been a major force that kept me going. Schoonpa and schoonma, I realize it is an immense privilege to have such warm, interested, and caring parents-in-law. Thank you for your support in the past years.

Finally, Aline, Bertin, Henriëtte, and Winston, you have been the best distraction from work one can ever imagine. Bertin, Henriëtte, and Winston, I know it is not the pedagogically correct thing to say, but you are the most wonderful children on earth. You are the future I have been working for. Aline, my deepest words of thanks go

to you. Thank you for believing in me and for your support, empathy, patience, and love. Thank you for everything.

Table of Contents

Acknowledgments **v**

1 Introduction **1**
1.1 Humanitarian Logistics 1
1.2 The Role of Evidence 2
1.3 Evidence-Based Humanitarian Decision Making 4
1.4 The Obstacles on the Road 6
1.5 Contributions 10
1.6 Overview of Chapters 12

2 Evidence-Based Vehicle Planning for Humanitarian Field Operations **15**
2.1 Introduction 15
2.2 Humanitarian Planning Context 17
2.2.1 General Description of Humanitarian Planning Systems 18
2.2.2 Current Practice 20
2.2.3 Cost-Effectiveness Framework 22
2.3 Model 27
2.3.1 Assumptions 27
2.3.2 Queuing Delay Distribution 28
2.3.3 Mean Field Delay 29
2.3.4 Disutility Functions 31

2.3.5 Expected Disutility 31
2.4 Numerical Analyses 32
2.4.1 Effectiveness of Systems that Pursue Proxy Objectives 33
2.4.2 Effectiveness of Data-Intensive Systems 35
2.4.3 Optimal Level of Centralization 37
2.4.4 Competing Planning System Adaptations 39
2.5 Conclusions and Discussion 41
2.A Differences with Private Sector Planning Problems 43
2.B Queuing System Analysis 44
2.C Mean Field Delay Function 48
2.D Interviewee Characteristics 50
2.E Questionnaires 50

3 The Roadside Healthcare Facility Location Problem 53
3.1 Introduction 53
3.2 Review of Related (Health) Facility Location Models 56
3.3 Continuous Access 58
3.3.1 Types of Service Packages 59
3.3.2 Access Measures 60
3.4 Model 62
3.4.1 Notation 63
3.4.2 Optimization Criteria 65
3.4.3 MIP Formulation 67
3.4.4 Worst-case Effect of Imprecise Input Data 69
3.5 Impact Analysis 70
3.5.1 South-East Africa Case 71
3.5.2 Network Expansion Strategies 75
3.5.3 Model Statistics 79
3.5.4 Sensitivity Analysis 80
3.5.5 Effect of the Number of New Facilities and Objective Weights. 81
3.6 Discussion and Conclusions 82
3.A Linearization of Effectiveness Variables e_{qs} 84

3.B Proof of Correctness of Constraints (3.17)-(3.21) 90
3.C Proof of Unique Solution for Polyhedron $P_B(\omega)$ 91
3.D Proof of Total Unimodularity of the Constraint Matrix Corresponding to Constraints (3.17)-(3.20) 92
3.E Proof of Strong $\mathcal{NP}$-hardness of RHFLP 92
3.F Proof of Proposition 3.2 94
3.G Proof of Proposition 3.3 94

4 A Column Generation Approach for Locating Roadside Clinics in Africa based on Effectiveness and Equity 97
4.1 Introduction 97
4.2 Literature Review 100
4.2.1 Equity 100
4.2.2 Facility Location and Access to Healthcare 102
4.3 Problem Description 103
4.3.1 Patient Volume 104
4.3.2 Continuous Access 105
4.3.3 Equity 107
4.4 Set Partitioning Type Formulation 110
4.5 Column Generation Approach 113
4.5.1 Restricted Master Problem 114
4.5.2 Pricing Problem 114
4.5.3 Initialization and Termination Criterion 115
4.5.4 Obtaining Integer Solutions 115
4.6 Acceleration Strategies 116
4.6.1 Accelerated Pricing 116
4.6.2 Adding Multiple Columns 119
4.6.3 Column Pool Management 119
4.6.4 Dual Stabilization 120
4.6.5 2-Stage Approach 120
4.7 Computational Experiments 120
4.7.1 Randomly Generated Instances 121

4.7.2 Integrality Gap . . . 122
4.7.3 Effect of Acceleration Strategies . . . 123
4.7.4 Overall Results . . . 125
4.8 Sensitivity Analysis on Equity . . . 127
4.9 Conclusions and Discussion . . . 130
4.A Direct Formulation . . . 131
4.B MIP Formulation Pricing Problem . . . 133
4.C Proof of Proposition 4.1 . . . 134
4.D Random Network Generation . . . 135

5 Forecasting Human African Trypanosomiasis Prevalences from Population Screening Data using Continuous Time Models 137
5.1 Introduction . . . 137
5.2 Materials and Methods . . . 140
5.2.1 Data . . . 140
5.2.2 Prediction Methods . . . 142
5.2.3 Model Fitting . . . 151
5.3 Results . . . 153
5.3.1 Fitted Models . . . 153
5.3.2 Comparison of Models . . . 155
5.3.3 Analysis of the Fixed Frequency Screening Policy . . . 157
5.4 Discussion and Conclusions . . . 163
5.A Notes Data Cleanup . . . 166
5.B Proofs on Screening Requirements for Eradication . . . 167
5.C Proofs on Screening Requirements for Elimination . . . 171
5.D Results for Maximum Likelihood Approaches . . . 173
5.E Sensitivity Analysis on Sensitivity and Participation Levels . . . 173
5.F Sensitivity Analysis on the Initial Value Assumption . . . 174
5.G Table of Notations . . . 175
5.H Prediction Errors . . . 176

6 Optimizing Population Screening for Infectious Diseases 177
6.1 Introduction . . . 177
6.2 Literature . . . 179
6.3 Problem Formulation . . . 183
6.4 Disease Burden Models . . . 185
6.5 Relationship between Capacity and Burden . . . 189
6.6 Solution Methods . . . 196
6.6.1 BLP Approach . . . 196
6.6.2 Iterated Local Optimization . . . 198
6.6.3 Column Generation Approach . . . 199
6.6.4 Planning Policies . . . 203
6.7 Case Study . . . 204
6.7.1 Baseline Case Description . . . 204
6.7.2 Computational Experiments . . . 207
6.7.3 Baseline Case Results . . . 208
6.7.4 Performance of Planning Policies . . . 210
6.7.5 End of Horizon Effects . . . 211
6.7.6 Sensitivity Analysis: Carrying Capacities . . . 211
6.7.7 Sensitivity Analysis: Screening Impact . . . 213
6.8 Conclusions and Discussion . . . 214
6.A Proofs . . . 216
6.B Upper Bounds on Screening Intervals . . . 224
6.C Repeated BLP Approach . . . 228
6.D Exact Solution Method Pricing Problem (6.39) . . . 229
6.E Local Search Method Pricing Problem (6.39) . . . 230
6.F Table of Notation . . . 230

7 Conclusions and summary 233

References 239

Nederlandse Samenvatting (Summary in Dutch) 265

Curriculum Vitae **271**

Portfolio **273**

Chapter 1

Introduction

1.1 Humanitarian Logistics

The Syrian civil war, Ebola, the earthquakes in Nepal and Haiti, the Japanese tsunami, the Indian Ocean tsunami, the ongoing HIV/AIDS epidemic,..., humanitarian caseload has been enormous in past decades. Between 2007 and 2016, disasters have affected two billion people, have killed 800,000 people, and caused 1.4 trillion dollars in damage (EM-DAT, 2017). Millions of people die from tuberculosis, HIV, or malaria every year (Murray et al., 2014) and one in ten people in the world lives below the poverty line of $1.90 per day (World Bank, 2016). Such humanitarian crises prompt substantial demands for humanitarian aid delivery. Billions of dollars are spent on disaster relief and development assistance every year. In 2015, for example, development assistance from governments alone amounted to 147 billion dollars (OECD, 2015).

Logistics plays a major and essential role in the humanitarian sector. Transportation is, after salaries, the largest cost category for international humanitarian organizations (Pedraza-Martinez et al., 2011). Moreover, the way logistics is organized has a direct impact on beneficiaries, for example, by affecting the speed and quantities of aid delivery. Optimization of this core activity has the potential to substantially aid beneficiaries by impacting both costs and effectiveness.

Optimization of logistics systems has long been at the core of the field of Operational Research and Management Science (OR/MS). Findings from this field, for example, assist hospitals in finding a feasible workforce schedule, railway operators in optimizing their network and timetabling, delivery companies in optimizing routes, manufacturers in optimizing their production process, warehouses in streamlining their order picking process, and supermarkets in determining optimal stock levels. The use of OR has enabled many companies to substantially reduce costs and at the same time improve service levels.

Despite the vast body of logistical knowledge and tools available, their uptake in the humanitarian sector has been lagging behind (Van Wassenhove, 2006; Gustavsson, 2003). This can be attributed largely to fundamental differences between typical commercial settings and humanitarian contexts. In contrast with its commercial counterpart, humanitarian logistics typically deals with security issues, high demand uncertainty, donor constraints, decentralized decision making, and challenging road and weather conditions (Pedraza-Martinez and Van Wassenhove, 2012; Holguín-Veras et al., 2012; Van Wassenhove, 2006). Moreover, whereas business logistics typically aims to minimize costs, humanitarian logistics typically aims to maximize (health) impact or minimize suffering. These differences hamper the transferability of commercial logistics knowledge and tools to humanitarian settings. In recent years, academics from the OR/MS community have started investigating the specifics of humanitarian contexts and how knowledge and tools from OR/MS could be of use in optimizing logistics in the humanitarian sector. This is the field of humanitarian logistics.

1.2 The Role of Evidence

Using the best available evidence to support humanitarian decision making is essential to maximize impact of scarce resources. Current practice in humanitarian organizations is, however, far from evidence-based (Banatvala and Zwi, 2000). Decision making is often undocumented and nontransparent, and evidence about "what works" in a given context plays a marginal role in decision making (Darcy et al., 2013). Instead, decisions are often determined by experience, instinct, and untested

assumptions. Though this is not necessarily bad, research shows that such decisions are prone to personal biases and disincentives and are very often determined by issues that have little to do with humanitarian objectives (Maxwell et al., 2013; Darcy et al., 2013).

Yet, attitudes seem to be changing. Recent initiatives like Evidenceaid, the Active Learning Network for Accountability and Performance (ALNAP), and the Humanitarian Evidence Program give proof of increasing awareness of the opportunities and the necessity of evidence-based humanitarian decision making. These programs synthesize and analyze available evidence, identify gaps, and promote evidence-based decision making in the sector. Humanitarian organizations themselves are also increasingly pushing for evidence-based decision making. This not only comes from an intrinsic motivation to maximize aid provided with scarce resources. Also donors are increasingly demanding cost-effectiveness. Organizations like GiveWell and The Centre for Effective Altruism, for example, evaluate and recommend charities based on their cost-effectiveness and have a vast impact on donor funding (Effective Altruism, 2017; GiveWell, 2017). For these reasons, humanitarian organizations increasingly push for incorporating evidence on (cost)-effectiveness in decision making (Knox-Clarke and Darcy, 2014). Marie Stopes International has an "Evidence to Action" team dedicated to "learn how to improve the effectiveness and cost-effectiveness of its programming" (Marie Stopes International, 2016b). Idem for World Vision International's "Evidence and Learning team" (World Vision International, 2017). Save the Children partners with a large network of universities to "build its evidence base on effective programming" (Save the Children, 2017). Similar initiatives exist in virtually every major humanitarian organization (cf. ICRC, 2015; OXFAM, 2015).

As humanitarian organizations become open to integrating evidence in their decision making, science could have a role in building the relevant evidence base. When humanitarian organizations are facing logistical challenges, the OR/MS community can provide answers. For example: Where to preposition relief goods? How much to preposition? How to set up a relief distribution network? How to operate vaccine supply chains? How to manage vehicle fleets? Where to position clinics? How to plan relief or development operations? What decision support systems to implement? What is the optimal degree of centralization of decisions?

Yet, providing clear to such questions answers is challenging. As we shall argue later, gathering empirical evidence is expensive and time-consuming, evidence resulting from a traditional modeling and optimization approach is often weak, and external validity of results is often limited due to large variations in contexts. Nevertheless, though substantial challenges must be overcome, OR/MS has the potential to make a substantial contribution to evidence-based humanitarian decision making. Before we elaborate on this, we first ask what evidence and evidence-based humanitarian decision making actually means and what makes certain types of evidence stronger than others.

1.3 Evidence-Based Humanitarian Decision Making

Evidence is commonly defined as "the available body of facts or information indicating whether a belief or proposition is true or valid"(McKean, 2005). In the context of humanitarian operations, two broad types of propositions can be distinguished (Knox-Clarke and Darcy, 2014):

- *Propositions about the existence of humanitarian needs resulting from a crisis*
- *Propositions about "what works" in addressing these needs*

Typical examples are the propositions that "Intervention is preferred to doing nothing" and that "A given intervention is better than alternative ones". In general, such propositions relate to specific *decisions* to be made in specific humanitarian contexts. In the field of OR/MS, we have traditionally approached such propositions from a modeling and optimization point of view. We make assumptions about the objectives to be reached, about suitable measures to proxy those objectives, about their relationship with decisions to be made, and about constraints on these decisions. This stylized version of the real life problem is then typically analyzed and solved using optimization techniques, and recommended actions and "management insights" are deduced from analytical results and case study analyses. Though there has been an increased interest in empirical studies in the OR/MS field, studies measuring

actual impact of proposed solutions, tools, or recommendations are still rare. Figures on articles published in the top 75 OR/MS journals are illustrative. Van de Klundert (2016) analyzed papers regarding diseases from the Global Burden of Disease top 15 diseases and found that only seven publications report on implementation of designed solutions and that only three report on health effects resulting from implementation.

Fields like health sciences followed a fundamentally different approach. Here, "the available body of facts or information" typically comes in the form of actual empirical results (Van de Klundert, 2016). For example, randomized trials and cohort studies are used to assess the proposition that "This intervention helps" or "This intervention is better than alternative ones". Moreover, medical research and practice are strongly driven by the notion that some forms of evidence are stronger than others; that is, certain types of evidence are more indicative of a proposition's validity than others. The Oxford Centre for Evidence-Based Medicine (CEBM) proposes hierarchies of study designs which can be used as "a short-cut for busy clinicians, researchers, or patients to find the likely best evidence" (Howick et al., 2011). To assess the proposition "This intervention helps", for example, they propose the following hierarchy:

1. *Systematic review of randomized trials or n-of-1 trials*
2. *Randomized trial or observational study with dramatic effect*
3. *Non-randomized controlled cohort/follow-up study*
4. *Case-series, case-control studies, or historically controlled studies*
5. *Mechanism-based reasoning*

The general approach of incorporating "best available evidence" in medical decision making has become deeply integrated in clinical practice in past decades. As defined by Sackett et al. (1996), it involves the "integration of clinical expertise, patient values, and the best research evidence into the decision making process for patient care". Note, hence, that evidence-based decision making does not mean that evidence *dictates* decisions. Contextual factors like patient values and patient char-

acteristics play a major role, and expertise is necessary to combine context with evidence.

A similar trend toward evidence-based decision making has occurred in the humanitarian sector. Several sets of criteria have been proposed to help decision makers in the humanitarian sector critically appraise and integrate external evidence (Knox-Clarke and Darcy, 2014). As in the medical sector, evidence does not dictate decisions. Substantial expertise is required to define the proposition (or question) at hand, to assess the context (objectives, constraints, resources, etc.), to search the evidence, and to interpret and combine this information so as to adequately assess the proposition's validity in the specific context considered. In parallel with the definition of evidence-based medicine (see Sackett et al., 1996), we define evidence-based humanitarian decision making as:

> *The integration of humanitarian decision makers' expertise, contextual information, and best available evidence into the decision making process for humanitarian operations.*

1.4 The Obstacles on the Road

We now return to the question whether and how OR/MS can contribute to building a useful evidence base for decision making in humanitarian contexts. According to the CEBM hierarchy of study designs, OR/MS studies typically produce the weakest type of evidence: we tend to use mechanism-based reasoning to relate decisions to outcomes, and rarely empirically analyze outcomes. This on the one hand constitutes a major weakness. Indeed, the field of OR/MS has many times been accused of producing and analyzing models that deviate too much from reality, leading to insights that are of little relevance to practice (Fisher, 2007; Corbett and Van Wassenhove, 1993). To strengthen the evidence base, implementation and evaluation of proposed solutions, tools, and recommendations will be key (cf. Gallien et al., 2015). Vledder et al. (2015) show that this is possible, using a randomized experiment (i.e., level 2 evidence) to evaluate different supply chain designs for essential medicines in Zambia.

On the other hand, we should not throw out the baby with the bathwater. The use of mechanism-based reasoning can provide "best available evidence" until higher level evidence is gathered (which may take years). Moreover, mechanism-based reasoning has at least three major advantages compared to empirical studies. First, it tends to be much cheaper and less time-consuming. Second, whereas empirical results on humanitarian interventions are hard to generalize due to large variations in contexts, mechanism-based reasoning enables one to explicitly model the role of contextual factors and to assess their impact on results. Third, whereas empirical studies are typically constrained to one or a few interventions, mechanism-based reasoning combined with optimization techniques enable one to select the "optimal" intervention. For these reasons, modeling and optimization studies can be of much value. However, in exploiting these advantages, modeling and optimization studies face at least five substantial challenges.

1. *Considering the broad scope of relevant objectives*

Evaluating proposed tools and solutions in terms of their effectiveness – e.g., through estimates of reductions in mortality or morbidity – is essential. Other criteria could may also be of major importance. Examples include costs, equity, humanitarian principles, and cultural and political acceptability (Bradt et al., 2009; Robertson et al., 2002; Banatvala and Zwi, 2000). Failure to incorporate such criteria may lead to poor recommendations. For example, advanced decision support tools can lead to effective decisions but can also be very inefficient due to large costs for implementation, maintenance, and training. Such tools may also be too time-intensive for situations requiring prompt decision making or may poorly fit a culture of decentralized decision making. In the same vein, sophisticated optimization techniques may be cost-effectiveness but may also fail in terms of required transparency (Vinck, 2013). Basing analyses on the full spectrum of relevant criteria is hence critical.

2. *Choosing objective functions that accurately reflect objectives*

A review of the humanitarian logistics literature reveals that very few papers take human suffering, which we can safely assume to be one of the major evaluation criteria of interest, as an explicit starting point for modeling choices. It is not uncommon to

come across models that minimize logistics costs rather than suffering and that weakly motivate why this is appropriate. Also the observation that suffering tends to increase convexly with response time (Holguín-Veras et al., 2013) is rarely incorporated (see, e.g., Holguín-Veras et al., 2013; De la Torre et al., 2012, for an overview of objective functions used for vehicle routing in disaster relief).

3. *Safeguarding internal validity of proposed models*

To adequately evaluate tools and solutions, models should accurately reflect reality. This poses at least two major challenges. The first challenge is to establish the link between decisions and objectives, i.e., the *mechanisms* linking the two. This link tends to be highly complex and context-specific, and empirical data required to quantify this link is oftentimes lacking (Banatvala and Zwi, 2000). For example, as we discuss in one of the chapters, it is hard to quantify how the location of clinics links to health outcomes through mechanisms like increased utilization, decreased treatment delay, and improved treatment adherence. Nevertheless, if OR/MS studies are to make a contribution on the basis of mechanism-based reasoning, it is important that assumed mechanisms are well-motivated and well-founded. The observation that very few modeling and optimization studies in humanitarian logistics make use of real data (Leiras et al., 2014) is therefore rather alarming.

The second challenge is to adequately model the decision space, which may be much smaller than expected. For example, security concerns, political issues, and donor constraints might leave little room for optimizing decisions. Feasible decisions can also be constrained by organizational culture. Our interviews with humanitarian organizations found, for example, that organizations are hesitant to centralize planning decisions, as this is perceived to reduce staff's professional freedom. When such issues are inadequately modeled, analyses of the (relative) effectiveness of proposed solutions or tools can be substantially biased.

4. *Analyzing external validity of results*

Diversity of humanitarian contexts makes effectiveness of proposed solutions and tools highly context-specific (see, e.g., Chapter 2). Generalizability of results is

therefore far from trivial. For example, centralized planning may be crucial in mass-casualty settings, but ineffective in more stable development settings. A given routing policy might work well for dense road networks but perform poorly when networks are sparse. Adequate assessment of the role of context is key to providing humanitarians with nuanced managerial insights. In specific cases, the link between context and results may be analytically expressed. In others, we have to rely on numerical analyses, e.g., based on a representative set of case studies or one case study combined with sensitivity analyses. Substantial progress is still to be made in that respect. A review of the humanitarian logistics literature by Leiras et al. (2014) found that only 23 of the 160 reviewed analytical papers included a case study.

Assumptions underlying numerical analyses also complicate generalizability. Numerical analyses commonly assume that data are available to the decision maker and that the decision maker can "optimally" set model parameters. To better assess the effectiveness of a proposed tool, one could question the impact of relaxing such assumptions. How is effectiveness affected when data are partly lacking or when decision makers make simplistic assumptions about them? What happens when the decision maker misinterprets a given model parameter?

5. *Choosing the right format*

The type of findings of a model-based study and the way they are presented will to a large extent determine their uptake by decision makers. Developing complex algorithms may be rewarding for the researcher, but simple guidelines likely have a much larger chance of being integrated in practice (cf. Knott, 1988). Similarly, closed-form mathematical expressions may be highly valued by fellow academics, but – if not translated into a jargon-free language – will likely make little sense to decision makers. Practitioners will rarely have time to read our analyses, which makes it important to present evidence in such a way that barriers to uptake are minimized (cf. Rousseau, 2006).

1.5 Contributions

The field of humanitarian logistics faces substantial challenges in contributing towards a useful evidence base for humanitarian decision making. This thesis does not present panaceas. On the contrary, critical analysis of the chapters to come will identify several areas where the strength of the evidence could be improved. Nevertheless, we do take several steps into the right direction.

First, the set of objectives considered in our models is based on extensive interactions with humanitarian organizations, including North Star Alliance, World Food Programme (WFP), United Nations High Commissioner for Refugees (UNHCR), Catholic Relief Services (CRS), GOAL Global, Plan International, and disease control program Programme National de Lutte contre la Trypanosomiase Humaine Africaine (PNLTHA).

Second, each of the presented chapters takes health outcomes or human suffering as an explicit starting point for modeling choices. Chapter 2 considers suffering or disutility, Chapters 3 and 4 focus on health outcomes, and Chapters 5 and 6 consider disease prevalence.

Third, the thesis presents two studies illustrating how "best available evidence" could be used to link decisions to outcomes, i.e., to establish and quantify the mechanisms linking the two. The first study, which is described in Chapters 3 and 4, considers network design decisions for roadside clinics that provide healthcare services to African truck drivers. To define the relationship between location decisions and health outcomes, we apply mechanism-based reasoning on the basis of expert interviews and medical literature. The second study, described in Chapters 5 and 6, considers the question how to deploy mobile teams that screen populations at risk of a given infectious disease. For the case of sleeping sickness control, Chapter 5 explicitly relates disease prevalence to the timing of screening rounds performed by mobile teams and thereby produces level 3 evidence: We retrospectively analyze this relationship for a cohort of villages. This yields an evidence-based objective function for Chapter 6, which considers how to optimize planning decisions for the mobile teams.

Fourth, our studies underline the importance of considering the full scope of relevant objectives. Chapters 2 and 6, for example, both stress the importance of accounting for the "costs of complexity", i.e., the costs of proposed tools in terms of implementation, training, and maintenance. We also address organizational culture as an important criterion, which is often overlooked in the literature. Additionally, we show that incorporating an extra objective does not need to come at a large cost in terms of current objectives. In our study on roadside clinics, for example, we show that solutions that are close to optimal in terms of each of the three relevant objectives are attainable. Similarly, we show that effectiveness of planning solutions is hardly affected when solely considering simple planning policies. In other words, tools that are close to "optimal" in terms of effectiveness and simplicity exist.

Fifth, the thesis contributes to awareness of the aforementioned challenges by showing the potential impact of misaligned objective functions and a lack of internal or external validity. Chapter 2, for example, discusses the importance of "context" in designing a planning system for humanitarian field operations. The "optimal" system depends highly on organizational, demand-related, and operational context factors, which makes generalizing (cost-)effectiveness claims potentially misleading. We also illustrate that data availability and data quality determine optimal planning system characteristics and hence ill-founded assumptions can result in poor recommendations. This shows the serious potential impact of a lack of internally valid models.

Finally, we extensively assess external validity of results. In some cases, we have been able to mathematically express the relationship between context and results. For example, Chapter 2 directly links effectiveness of planning decisions to factors like road density, travel time variability, and urgency. Chapter 5 mathematically expresses screening efforts required to eradicate or eliminate the considered disease as a function of factors like participation levels, sensitivity of diagnostics, and endemicity. In other chapters, we rely on numerical sensitivity analyses. For example, Chapters 3 numerically analyzes how effectiveness of location decisions changes with data quality and with the user's capacity to adequately set model parameters. Similarly, Chapter 6 analyzes the impact of estimation errors on the effectiveness of planning policies.

1.6 Overview of Chapters

Below, we briefly summarize contributions of each chapter. Apart from minor adaptations, the chapters are exact copies of papers submitted to or published in academic journals. The majority of the work in Chapters 2, 3, 5, and 6 has been done independently under close supervision of co-authors. The main contributions of the work in Chapter 4 were to generate the idea, develop the case study, and write the paper. The first author's main contributions were to generate the idea, implement the methods, and run the experiments.

Chapter 2: H. de Vries, L.N. van Wassenhove. Evidence-Based Vehicle Planning for Humanitarian Field Operations. Submitted to Manufacturing & Service Operations Management.
This chapter discusses the applicability and cost-effectiveness of advanced approaches to optimize planning and routing of humanitarian field operations. Combining insights from expert interviews, literature, and extensive numerical analyses, we show that "optimal" planning system characteristics depend strongly on organizational, demand-related, and operational context factors.

Chapter 3: H. de Vries, J.J. van de Klundert, A.P.M. Wagelmans. The Roadside Healthcare Facility Location Problem. Submitted to Production and Operations Management. This paper won the INFORMS Healthcare 2015 best student paper award.
In this chapter, we consider the problem of selecting locations for clinics along African highways so as to maximize their impact in terms of truck driver patient volume and health service effectiveness. We develop models for this problem, present numerical experiments for the network of major transport corridors in Southern and Eastern Africa, and discuss policy implications.

Chapter 4: J. Nunez Ares, H. de Vries, D. Huisman. A Column Generation Approach for Locating Roadside Clinics in Africa based on Effectiveness and Equity. Published in European Journal of Operations Research.
This chapter analyzes the trade-off between effectiveness and equity of healthcare delivery among African truck drivers. We show that networks that are close to optimal with respect to both objectives can be designed efficiently through column generation techniques.

Chapter 5: H. de Vries, A.P.M. Wagelmans, E. Hasker, C. Lumbala, P. Lutumba, S.J. de Vlas, J.J. van de Klundert. Forecasting Human African Trypanosomiasis Prevalences from Population Screening Data using Continuous Time Models. Published in PLoS Computational Biology.
In this chapter, we propose several models for the relationship between the timing of screening operations for Human African Trypanosomiasis (HAT, also called sleeping sickness) and disease prevalence. The models are then fitted and tested by means of a dataset with data on prevalence levels and screening operations in the Democratic Republic of Congo (DRC) and used to analyze screening requirements for elimination and eradication.

Chapter 6: H. de Vries, A.P.M. Wagelmans, J.J. van de Klundert. Optimizing Population Screening for Infectious Diseases. In preparation for journal submission.
The final chapter considers the problem of planning population screening operations so as to minimize the burden of infectious disease. For a broad class of diseases, we propose general solution methods and simple planning policies, as well as general methods to analyze the relationship between screening capacity and long term disease burden. Using the models developed for HAT in Chapter 5, we show that the planning policies recommended by the WHO could likely be improved upon and that simple policies are hardly inferior to approaches employing advanced optimization techniques.

Finally, Chapter 7 summarizes the main findings presented in this thesis.

Chapter 2

Evidence-Based Vehicle Planning for Humanitarian Field Operations[1]

2.1 Introduction

Decision support software has substantially transformed private sector logistics and enabled companies to significantly decrease transportation costs and/or improve service levels. This has not happened yet in humanitarian logistics. Given the substantial funds spent on humanitarian operations, this is rather surprising, especially since there has been a significant increase in humanitarian case load. Logistics efforts play a key role in delivering disaster relief and development services and transportation is, after salaries, the largest cost category for international humanitarian organizations (IHOs) (Pedraza-Martinez et al., 2011). The cost of 4x4 vehicle fleets of IHOs is estimated to exceed $1 billion per year and the number of vehicles is expected to triple

[1]Apart from several minor adaptations, this chapter is a direct copy of the article: H. de Vries, L.N. van Wassenhove (2017). *Evidence-Based Vehicle Planning for Humanitarian Field Operations*. Submitted to Manufacturing & Service Operations Management.

by 2050 (Pedraza-Martinez et al., 2011). One would expect that advanced planning and routing software could have a substantial positive impact.

Gustavsson (2003) suggests three internal hurdles that have kept IHOs from realizing these apparent gains: (1) lack of logistics expertise, (2) undervaluation of IT systems, and (3) difficulties in getting the necessary funding. Literature and semi-structured interviews with fleet managers and logistics experts from IHOs reveal three additional hypotheses:

1. *The effectiveness increase realized by advanced planning and routing does not outweigh the cost of implementation, operation, and maintenance.*

2. *Available planning solutions do not fit the complex context of humanitarian organizations. E.g., they may be too data-intensive or pursue objectives not aligned with humanitarian missions.*

3. *Implementing advanced planning and routing competes with innovations and responsibilities that have a higher priority.*

This chapter aims to shed a light on the cost-effectiveness of advanced planning and routing approaches in humanitarian contexts. In Section 2.2, we develop a framework for cost-effectiveness based on our literature review and a set of interviews with humanitarian logistics experts. Next, Section 2.3 proposes a simple model that relates operational effectiveness to the specific context of IHOs and to their planning systems. This model then serves as a basis for extensive numerical analyses in Section 2.4.

In general, our results urge researchers and decision makers *to adopt an evidence-based approach when evaluating planning and routing software* by considering external validity of results, internal validity of underlying models, and by extending evaluation criteria beyond effectiveness alone. More specifically, we make seven contributions. First, we fill the gap in literature on the cost-effectiveness of advanced planning and routing systems in IHOs. Second, we add nuance to the literature on humanitarian last-mile logistics by highlighting the types of costs and counter-productivities that may come with advanced planning approaches. Third, we propose a simple and

intuitive model for the effectiveness of a planning system, which could serve as a basis for further modeling studies on fleet management in humanitarian organizations. Fourth, we provide evidence stressing the importance of a careful analysis of the operational context when choosing what planning system to implement. Fifth, we show how (in)effective standard planning and routing solutions can be in the humanitarian context. Sixth, we hypothesize general relationships between contextual factors and the optimal planning system. Finally, we assist decision makers in prioritizing adaptations of their planning system (e.g., stimulating car pooling and increasing mission lengths) by showing their respective potential impact.

2.2 Humanitarian Planning Context

This section provides background knowledge on the terminology used to define a humanitarian planning system (Section 2.2.1), current planning practices (Section 2.2.2), and determinants of the cost-effectiveness of a planning system (Section 2.2.3). Our results are based on a literature review and in-depth interviews with eight logistics experts from a representative set of humanitarian organizations. These include the World Food Programme (WFP), United Nations High Commissioner for Refugees (UNHCR), Catholic Relief Services (CRS), GOAL Global, Plan International, and disease control program Programme National de Lutte contre la Trypanosomiase Humaine Africaine (PNLTHA). The interviews (see Appendix 2.E) were conducted by phone or Skype and took on average 40 minutes. The aim was to reveal current planning practices, inefficiencies therein, and the potential role of decision support tools. We also interviewed three consultants from the IT company ORTEC Consulting who had been involved with software applications for the humanitarian sector.

To avoid confusion on terminology, we define that the *planning* of humanitarian operations refers to deciding on the (approximate) timing of the delivery of goods and services to beneficiaries. We refer to the latter as *operations*. *Routing* is concerned with determining the actual routes to be taken by mobile units. A *planning system* defines the procedures and methods used to take planning and routing decisions. With *advanced planning and routing* we refer to mathematically optimized planning and routing based on detailed information on travel times and requests. We define

operational *effectiveness* as the extent to which the IHOs' objectives are reached, as partly determined by the planning system. We assume that these goals are to minimize the harm, suffering, health burden, distress, or inconvenience caused by humanitarian crises (cf. Holguín-Veras et al., 2012), which we jointly refer to as *disutility*. Disutility could be measured in terms of (deprivation) costs or health outcomes (cf. Holguín-Veras et al., 2013; Yadav, 2010). When analyzing *cost-effectiveness* of planning systems, we compare them in terms of costs and effectiveness.

2.2.1 General Description of Humanitarian Planning Systems

A planning system can be defined in terms of the following characteristics: (1) centralized, decentralized, or hybrid, (2) periodic review vs. continuous review, (3) one vs. multiple destinations per trip, (4) pooled vs. dedicated, (5) impact-based vs. proxy metric-based. In a *centralized* system, a person, a team, or an IT system recommends or makes decisions, whereas local staff or a driver makes these decisions in a *decentralized* system. *Hybrid* systems combine the two structures, for example by making centralized planning decisions and by allowing local staff to determine the routing. In contrast with a *continuous review* system, a *periodic review* system generates planning and routing decisions at fixed time intervals (e.g., weekly). When vehicles or other transportation means visit *one destination* per trip, resource allocation and prioritization are the major planning decisions, whereas routing becomes relevant in the *multiple destinations* case. In *pooled systems*, every vehicle can be assigned to a vehicle request, whereas vehicles are dedicated to a specific category of requests in the *dedicated system*. Finally, when planning and routing is optimized in terms of the disutility the decisions avert, we refer to this as *impact-based* planning and routing. Systems that optimize with respect to other metrics like travel times are referred to as *proxy metric-based*.

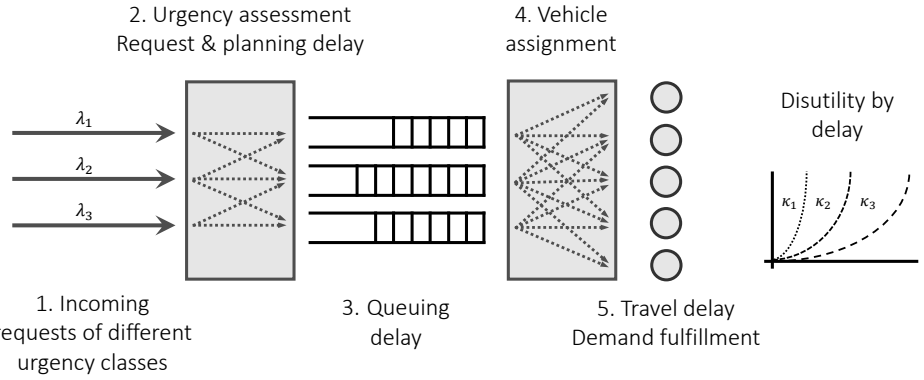


Figure 2.1: The flow of a vehicle request through the planning system.

Figure 2.1 depicts how a given planning system, as defined by these five characteristics, relates to disutility, and describes the general context considered in this work. We assume a disaster relief or development program that has a set of vehicles at its disposal which it can use to perform its operations. Requests for vehicles evolve according to a stochastic process, which occurs at (1) in Figure 2.1. Without loss of generality, we assume that one request relates to one destination. Requests vary in terms of their urgency levels (e.g., high, medium, and low). Fulfillment of these requests may incur *request delay* at (2) when requesters do not immediately reveal them and *planning delay* in case of a periodical planning system. Depending on the planning system, requests may be assessed in terms of their urgency levels and may be assigned a priority accordingly. This assessment may be prone to errors. If all vehicles are in use, the request cannot immediately be satisfied and incurs what we refer to as *queuing delay* at (3). As soon as a vehicle becomes available, it will be assigned to a new request in queue. Note that a vehicle may combine multiple requests in one route, but that we consider a request to be in queue as long as the vehicle is visiting preceding destinations. Assignment of vehicles to requests at (4) is determined by the planning system, which may take priority levels and travel times as inputs, and may be constrained by earmarking constraints, by a lack of car pooling

among different units or programs, and by vehicle equipment requirements. The time it takes to meet a request at (5) is referred to as *field delay*. It consists of the time it takes to travel to the corresponding destination, which we refer to as *travel delay*, and the time spent "on site". Both may be highly stochastic. Travel delay depends on the planning and routing system and contextual factors like security issues, weather, and road conditions. In line with its definition, effectiveness is measured as the expected disutility incurred *by a random request* as a consequence of delaying its fulfillment. The total delay incurred by a request equals the sum of request delay, planning delay, queuing delay, and field delay. The relationship between disutility and total delay is quantified by a disutility function, and differences in urgency levels are reflected by differences in disutility functions.

2.2.2 Current Practice

Planning and routing practices vary widely between organizations and even between different programs within the same IHO (see, e.g., Pedraza-Martinez et al., 2011; UNHCR, 2006). This can (at least partly) be explained by a lack of guidelines and/or resources to implement them.

> *"Fleet management is largely left to the regions and the regions leave it to the countries. If there happens to be someone with some expertise in fleet management then this is fine, otherwise there happens very little to reach the full potential."* [Respondent 1]

> *"Organizations know how to do fleet management. We have the guidelines and the manuals. Yet, budget determines the space you get to implement structures."* [Respondent 2]

Next, without trying to generalize, we will mention some general tendencies distilled from our interviews and literature review. In many programs, particularly in a disaster relief context, planning is very much decentralized (Holguín-Veras et al., 2012). The decision maker simply gets transportation requests from field staff and tries to meet them as soon as possible (cf. Pedraza-Martinez et al., 2010).

"Based on when they come in, they will just be assigned a vehicle." [Respondent 3]

Requests are typically communicated verbally in smaller programs and via e-mail or request forms in larger programs. The decision maker can be a fleet manager, a planner or dispatcher, or a team involving field and/or support staff. In some cases, support staff proposes a planning which is then sent for approval to the program office. In other cases, the planner or planning team itself proposes planning and/or routing decisions. In a context where urgency is relatively low, this is typically done on a weekly basis, and updated if needed (Pedraza-Martinez et al., 2011). None of the organizations involved was reported to use IT tools to support these decisions, even though some have access to vehicle booking functionalities available in fleet management software. Costs and staff capacity are mentioned as major reasons, resonating with our first hypothesis (see Section 2.1). Instead, many organizations use simple spreadsheets and whiteboards.

"It is a matter of staff capacity: 1) not enough staff, and 2) staff that is not sufficiently skilled to use this." [Respondent 1]

"It is a matter of cost-benefits. Such IT system may deliver slightly higher efficiency, but may cost one vehicle." [Respondent 2]

"They have a whiteboard, they (...) put in their requests by Friday afternoon, they establish (...) the next seven day planning, and then they e-mail that out to all staff at the end of the day." [Respondent 4]

Urgency of requests is an important planning criterion (see, e.g., Greaney et al., 2011; Waldman, 2001) and is typically communicated to the dispatcher through a verbal description or through (color) code tags (see, e.g., Lerner et al., 2008). Freedom to plan activities in line with this main criterion is, however, sometimes limited by requests that have to be fulfilled at specific times (e.g., to attend a meeting) or by security issues. For example, one interviewee mentioned that planning can be largely determined by the availability of military convoys. Decision space can be further affected by earmarked funding (Besiou et al., 2012). Donor constraints can refrain

programs from car pooling and fleet sharing between programs, causing them to operate in silos:

> *"Some projects are covered by donors for 100%. The vehicles are then dedicated to that project only. Donors are very strict on that. They may even check this in their audits."* [Respondent 1]

Vehicle use can also be constrained *within* programs, for example by dedicating vehicles to specific units.

> *"(...) finance will have their own vehicles, programs will have their own vehicles, logistics will have their own vehicles."* [Respondent 5]

Planning practices are largely determined by the specifics of the *humanitarian context*. For example, time pressure and lack of information heavily constrain informed and centralized decision making in disaster relief contexts (Holguín-Veras et al., 2012). The *type of aid* delivered also largely affects planning needs and planning practices. In the context of basic relief item delivery, point-to-point (PTP) or full truckload (FTL) transportation is very common (Pérez-Rodríguez and Holguín-Veras, 2015), rendering routing optimization trivial. In contexts of less than full truckload (LTL) transportation and service delivery, routing *can* constitute major opportunities for optimization. For example, NGO Marie Stopes International substantially saves travel times through routing optimization for its outreach teams (Marie Stopes International, 2016a). *Program size* is a third major determinant. In smaller programs, the planner tends to have several other responsibilities, which affects the planning system.

> *"When you have 10 responsibilities and fleet management is one of them, you only focus on the very basics."* [Respondent 1]

2.2.3 Cost-Effectiveness Framework

We now return to the main goal: to assess cost-effectiveness of advanced planning and routing approaches. As can partly be seen from Figure 2.1, cost-effectiveness

can be affected in three ways: through a direct or indirect impact on (1) costs, (2) delay, and/or (3) the relationship between delay and disutility. Next, we combine insights obtained from our interviews and literature review to design a framework of determinants impacting these three factors. Determinants are classified as organizational, planning system, operational context, and demand-related factors, and are depicted in Figure 2.2. We assume all costs and resources to be constant, except for those related to the planning system.

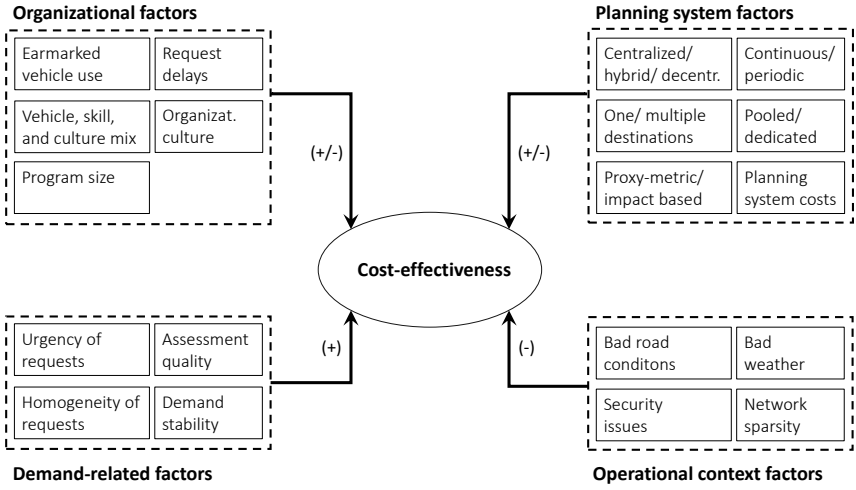


Figure 2.2: Contextual framework for operational cost-effectiveness.

Organizational factors. Program staff tends to reveal its transportation needs at the last moment, although often they could have anticipated them much earlier (Pedraza-Martinez et al., 2011). These request delays create an "artificial demand uncertainty", which affects cost-effectiveness. A second determinant is the level of earmarked vehicle use (Besiou et al., 2012). This can lead to a dedicated planning system with some disadvantages. Effectiveness is also affected when specific teams must fulfill certain requests as opposed to any team being able to fulfill any request. For example, ethnic violence made one organization stop sending drivers from certain

tribes to Darfur. Vehicle characteristics, like the age of a vehicle or specific equipment, may also induce assignment constraints (Eftekhar and Van Wassenhove, 2016). Organizational culture is a fourth determinant, as it determines the acceptability and thereby the effectiveness of a planning system. We return to this issue later. Finally, program size determines *relative* cost-effectiveness of different approaches, as we discuss later.

Planning system factors. Cost-effectiveness is affected by each of the five characteristics of planning systems. First, *decentralized* systems can result in a lack of coordination (cf. Pedraza-Martinez and Van Wassenhove, 2012; Stapleton et al., 2009; UNHCR, 2006). As a potential consequence, urgency levels are not incorporated appropriately in routing and prioritization decisions. Planners in *centralized* systems may lack (reliable) contextual information. Indeed, IT systems and dispatchers may have relatively little insight into security issues, weather and road conditions, demand mobility and their effects on travel times (Pedraza-Martinez and Van Wassenhove, 2012; UNHCR, 2006).

> *"There is no single software that knows the roads better than us. We go to the places where nobody else wants to go. You won't find these roads on Google."* [Respondent 6]

In case of such information gaps, *hybrid systems* may be more effective, e.g. by combining priority setting on a centralized level with autonomy in planning and routing on a local level. Second, a *periodic review* system has the advantage that the pool of destinations to be planned tends to be larger, yielding more efficient routes (cf. Daganzo, 1984). Planning delays induced by such system form a disadvantage. Third, visiting more than *one destination* in one trip tends to decrease travel and queuing delays (cf. Daganzo, 1984). Fourth, systems in which vehicles are *dedicated* can block the assignment of available vehicles to requests, and thereby induce queuing delays. Moreover, pools of destinations will be smaller, which increases travel delays. Finally, *impact-based planning* has an obvious effectiveness advantage over proxy metric-based planning, but may bring about substantial additional costs in terms of urgency assessments and implementation and maintenance of solution methods.

This brings us to a general major determinant of cost-effectiveness: the costs of a planning system. Such a system may require an IT solution, possibly including expensive vehicle routing software, and a planner or dispatcher. Moreover, implementing (rolling out) such a system will require substantial amounts of training and consume scarce human resources and budgets (cf. Winters et al., 2008). IT systems typically also require expensive support and maintenance services (Koskinen, 2010). Finally, planning may require time-consuming activities such as data gathering, information exchange (planner vs. staff and drivers), and urgency assessments.

Operational context factors. Travel times are determined by the road network, security issues, weather, and humanitarian issues (e.g., disasters might destroy roads and bridges). These issues have an obvious impact on cost-effectiveness as they determine travel and queuing delays and whether or not a destination can be visited at all. Sparsity also has an impact on the *relative* cost-effectiveness of advanced planning systems compared to simpler ones, as we discuss later. The same holds in general for stochasticity in travel times, as caused by varying operating conditions.

Demand-related factors. Urgency levels of requests constitute a major determinant of cost-effectiveness. Higher urgency levels imply that larger amounts of disutility can potentially be averted and hence an improved cost-effectiveness could be attained.

Quality of the urgency assessment and possible communication noises determine to what extent the planning decisions reflect the relative urgency levels of requests, and hence the disutility averted (see, e.g., Frykberg, 2005). Assessment quality may also be affected by misaligned incentives. For example, one interviewee reported that teams sometimes misrepresent needs in order to visit girlfriends or to reach certain targets.

Homogeneity of requests in terms of the required skills, equipment, and/or cultural background of staff is another demand-related factor. Heterogeneity can induce vehicle assignment constraints and hence queuing and travel delays. Finally, cost-effectiveness is also affected by demand mobility. For example, in case of conflicts,

demand locations and demand sizes are difficult to predict (Holguín-Veras et al., 2012), which hinders effective resource allocation and prioritization.

Interaction effects. The extent to which a given factor affects cost-effectiveness can be affected by interaction effects. First, the amount of variation in urgency levels determines the relevance of prioritization and hence the relative effectiveness of impact-based planning. Second, the amount of operational uncertainty will be highly correlated with the size of information gaps on a central level and hence with the relative effectiveness of centralized planning systems. Real-time information systems are virtually absent in the humanitarian context and much local information is not being institutionalized (Pedraza-Martinez and Van Wassenhove, 2012; UNHCR, 2006). Third, the number of possible routes tend to be smaller in sparser networks, rendering the corresponding optimization problems easier and hence the relative effectiveness of advanced planning approaches compared to simpler ones smaller:

> *"In most operations, on the local level, there is normally only one route, which can either be closed or open due to the security situation. The planning options on using different routes are very few."* [Respondent 2]

Similarly, security issues affect relative effectiveness when they require routes to be unpredictable rather than efficient.

> *"(...) there are often kidnappings of the UN workers in the area. One has to change and you should not be predictable in order to survive (...)."* [Respondent 6]

Fourth, program size and planning system factors have a major joint effect on cost-effectiveness. Investments needed to install and run an advanced planning system may, for example, have a relatively smaller impact on cost-effectiveness when program size grows. Fifth, request homogeneity interacts with vehicle, skill, and cultural mix. These factors jointly determine the space of feasible assignments and hence what queuing and travel delays will be incurred. Finally some planning systems may better fit the culture of a humanitarian organization than others. Systems involving black box optimization, for example, may fail to meet humanitarian standards concerning

transparency (Vinck, 2013). Moreover, having a dispatcher telling field staff what to do and where to go may cause frustrations and discrepancies between what staff perceives as being needed and what is requested. As a result, such systems may fail.

> *"You cannot expect staff to simply accept [planning] recommendations."* [Respondent 1]

> *"When you tell them "this is not efficient", they get angry (...)."* [Respondent 7]

> *"Planning is about culture."* [Respondent 4]

More generally, the system affects the amount of autonomy staff has and the amount of bureaucracy they encounter. Studies among social workers show that these two factors have a substantial impact on job satisfaction, the risk of getting a burnout, and staff turnover (Kim and Stoner, 2008; Arches, 1991), each of which may clearly affect effectiveness.

2.3 Model

As explained, we are interested in the cost-effectiveness of advanced planning and routing systems in different humanitarian contexts. Since costs will be highly context- and organization-specific and are sometimes hard to quantify, we shall focus our analyses on context-specific *effectiveness*. Decision makers can then weigh the results against the specific cost structures they encounter.

The aim of our analyses is to provide general insights into how effectiveness relates to planning system characteristics and contextual factors and to indicate the order or magnitude of effectiveness gains that could be achieved when implementing one system instead of another. In line with this purpose, we present a simple model that enables us to obtain results in an exact manner.

2.3.1 Assumptions

Our analyses assume the general request flow depicted in Figure 2.1. Requests are classified into two classes, those of higher urgency (H) and those of lower urgency (L),

and may incur request delay, planning delay, queuing delay, and field delay. Delays are largely affected by the planning system. For a given system, let us consider a random request of urgency class $i \in I = \{H, L\}$. We denote the four types of delay incurred by *random variables* X_i^R, X_i^P, X_i^Q, and X_i^F and the total delay by random variable X_i. The latter is characterized by some probability density function $f_{X_i}(x_i)$. Furthermore, function $u_i(x_i)$ quantifies the total expected disutility a class i request incurs when its total delay equals x_i time units. Combining the probability function and the disutility function yields the total expected disutility incurred by a random request in the system, which we use as an indicator of effectiveness:

$$\mathcal{U} = \pi_H \int_0^\infty f_{X_H}(x_H) u_H(x_H) dx_H + \pi_L \int_0^\infty f_{X_L}(x_L) u_L(x_L) dx_L \tag{2.1}$$

Here, π_i denotes the probability that a request belongs to urgency class i. In what follows, we develop an explicit expression for $\mathcal{U}$ based on the following simplifying assumptions:

Assumption 2.1. *Class i requests evolve according to a Poisson process with rate λ_i*

Assumption 2.2. *The sum of request delay and planning delay, $X_i^{RP} = X_i^R + X_i^P$, is exponentially distributed with parameter $\lambda^{RP} = \lambda^R + \lambda^P$*

Assumption 2.3. *Field delay is exponentially distributed with parameter $\lambda^F = \mu^T + \mu^S$*

Assumption 2.4. *Variables X_i^{RP}, X_i^Q, and X_i^F are independent*

Here, parameters λ^R, λ^P, μ^T, and μ^S denote the reciprocal of the mean request delay, planning delay, travel delay, and time on site, respectively.

2.3.2 Queuing Delay Distribution

We split the planning system into a part in which request and planning delays are incurred (part 1) and a part in which queuing delays and field delays are incurred

(part 2). The system can be seen as a simple queuing network in which servers represent vehicles and clients represent requests. Due to Assumptions 2.1 and 2.2, part 1 becomes an $M/M/\infty$ queuing system where requests simply incur an exponentially distributed service time.

Let us consider the case where requests are assigned to vehicles on a first-come-first-served (FCFS) basis. Because of Assumption 2.3 and because the arrival process for part 2 is Poisson (Mirasol, 1963), this part becomes an $M/M/C$ queuing system, in which C denotes the number of vehicles. Queuing theory states that X_i^Q is exponentially distributed with some parameter λ^Q given that an arriving request has to wait (because all servers are busy). The latter occurs with some probability Π. Expressions for Π and λ^Q are given in Appendix 2.B.

As an alternative, one could first assess the urgency of an arriving request and then assign it to a priority queue. Requests assigned to a high priority queue then have priority over those assigned to the low priority queue when a vehicle becomes available. Let q_{ij} denote the probability that a class i request ends up in queue $j \in J = \{H, L\}$. This parameter hence measures the quality of urgency assessments. The probability that a request is of urgency class i *and* ends up in queue j is denoted by π_{ij} and equals $\pi_i q_{ij}$. Part 2 then represents an $M/M/C$ *priority queuing system*. Due to the analytical complexity of the exact queuing delay distribution (see Davis, 1966), we propose a simple but close approximation by making the following simplifying assumption:

Assumption 2.5. *Variable X_i^Q equals 0 with probability $(1-\Pi)$ and is exponentially distributed with parameter λ^{Q_j} with probability $q_{ij}\Pi$.*

Though this assumption does not misrepresent mean delay and delay probability Π, it does slightly misrepresent the delay distributions. Expressions for Π and λ^{Q_j} are provided in Appendix 2.B, as is an analysis of the quality of the approximation.

2.3.3 Mean Field Delay

Recall that field delay is the sum of travel delay and time on site. We assume that the latter is not affected by the planning system, and hence that μ^S represents some

context-specific constant. Expected travel delay for any request depends on system characteristics and contextual factors like the road network. To keep our results general, we estimate *expected* delays for networks with certain properties rather than delays for one specific road network. Specifically, we analyze expected delays per request in random graphs in the square $[0, \tau]^2$. Here, τ denotes a scaling factor such that expected travel times equal Euclidian distances. Let N denote the number of destinations to be visited, K the number of destinations per trip, $\eta \in [0, 1]$ some measure of network sparsity (see Appendix 2.C), and $\delta \in [0, 1]$ the level of travel time stochasticity. Regarding the latter, we assume that the real travel time $\tilde{t}_{ij}$ for arc (i, j) deviates up to a fraction δ from the expected travel time, i.e. Euclidian distance, t_{ij} according to a uniform distribution. Finally, φ represents the specific routing procedure used. Appendix 2.C shows that $T_\varphi(N, K, \eta, \delta)$ – the expected travel delay per request (i.e., $\frac{1}{\mu^T}$), given the indicated parameters – is well estimated by the following formula:

$$T_\varphi(N, K, \eta, \delta) \approx T_\varphi(N, K, 1, 0) \left(1 + \beta_{\varphi 1}(1 - \eta)^{\beta_{\varphi 2}}\right) \left(1 - \beta_{\varphi 3}\delta^{\beta_{\varphi 4}}\right) \tag{2.2}$$

Here, $\beta_{\varphi 1}, \beta_{\varphi 2}, ...$ are some constants and $T_\varphi(N, K, 1, 0)$ the expected travel delay in a full graph. We shall fit this function for three routing procedures. The first represents the case where routing decisions are only constrained by the number of destinations per route. In the second, routing must also adhere to precedence constraints. Specifically, as many vehicles as possible must visit at least one high priority destination and each vehicle must first visit all high priority destinations assigned to it. (Recall that a destination corresponds to a high or low priority request.) The third represents the case where routing decisions must also adhere to clustering constraints. Specifically, we first divide the region into four quadrants and then optimize routing within each cluster while adhering to the aforementioned precedence and capacity constraints. We refer to the three vehicle routing problem (VRP) variants corresponding to these procedures as the CVRP, the PC-CVRP, and the CL-PC-CVRP. Acronym CVRP stands for the "capacitated vehicle routing problem" and acronyms PC and CL stand for "precedence constrained" and "clustered", respectively.

2.3.4 Disutility Functions

Holguín-Veras et al. (2013) argue that disutility (which they refer to as "deprivation costs") is expected to be a monotonic, non-linear, and convex function of delay. Moreover, they show that the willingness to pay for water when the time without water equals t is well described by a function of the form:

$$u(t) = \gamma(e^{\kappa t} - 1) \tag{2.3}$$

Here, γ and κ denote some strictly positive constants. The authors also propose several proxy functions including fixed, variable, and infinite penalty functions. Alternative disutility functions may include an upper bound on disutility. One example is the logistic (s-shaped) function, which is commonly used to measure the evolution of the disease burden resulting from an epidemic (see, e.g., De Vries et al., 2016).

2.3.5 Expected Disutility

Figure 2.3 summarizes our assumptions about the delay distributions. In Appendix 2.B, we prove a simple, closed-form expression for the expected disutility incurred by a random request under the additional assumption that disutility is quantified by function (2.3). The expression takes the following as input: (1) a parametrization of the disutility function, (2) the mean field delay, (3) the mean planning and request delay, (4) the probability that a class i request ends up in queue j, (5) the probability that queuing delay is incurred, and (6) the conditional average queuing delay. Here, for a given routing procedure, mean field delay is a simple function of network sparsity, travel time stochasticity, the number of destinations, the number of destinations per trip, and a travel time scaling factor τ. Furthermore, (5) and (6) are simple functions of the rate at which class i requests evolve, the number of vehicles, and the mean field delay.

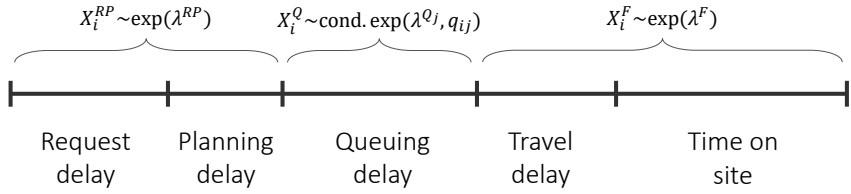


Figure 2.3: Model assumptions about the distribution of delays incurred between generation and fulfillment of a request.

2.4 Numerical Analyses

We now use the closed-form expression to compare alternative planning systems. Throughout this section, we make use of a "baseline case" which is based on a realistic humanitarian context. Our parameter estimates, presented in Table 2.1, were obtained through a questionnaire and a subsequent interview with a logistics expert from one of the humanitarian organizations involved in this study. The questionnaire asks for several key metrics describing a "typical program", and can be found in Appendix 2.E.

Parameter	Value	Remark
τ	0.544 days	Calibrated
η	0.0128	
δ	0.25	
N	40	Assumption
K	2	
C	22	Assumption
λ_H, λ_L	10 per day, 10 per day	
q_{ij}	0.75 if $i = j$, 0.25 otherwise	
κ_H, κ_L	1/10, 5/70	Not unique
γ_H, γ_L	1, 1	Not unique
λ^R, λ^P	1/1.5 per day, 1/2 per day	
μ^S	1/0.648 per day	Calibrated

Table 2.1: Baseline case parameters.

Our questionnaire results do not yield unique disutility functions. Instead, as suffices for the purpose of our analyses, the questionnaire aims to reveal a realistic quantification of the *relative disutility* of higher and lower urgency requests. Parameter τ was chosen such that the average travel time from depot to destination equals the reported value. Mean time on site was estimated by deducting the estimated travel time per request from the reported field delay per request. The reported number of vehicles C was slightly adjusted to obtain a stable queuing system. Finally, our experiments involve solving thousands of VRP instances, for which computation times grow exponentially with N. We therefore restricted ourselves to instances involving up to 40 nodes (i.e., demand points).

The functions $T_\varphi(N, K, \eta, \delta)$ we use in our analyses are defined in Table 2.2 and were obtained through least squares fitting (see Appendix 2.C). VRP problem instances were solved by applying CPLEX 12.63 to the two-commodity flow formulation from Baldacci et al. (2004) combined with rounded capacity inequalities from Lysgaard (2003) and (if applicable) precedence constraints. The solver was programmed in C++ and validated against standard instances from Christofides and Eilon (1969). Our overall experiments were programmed in Matlab R2015a.

Parameter	Value	# Instances	R^2
β_{CVRP}	[0.782; 21.263; 0.336; 1.451]	12100	0.98
$\beta_{PC-CVRP}$	[0.809; 21.271; 0.172; 1.803]	12100	0.95
$\beta_{CL-PC-CVRP}$	[0.840; 20.812; 0.159; 1.607]	12100	0.94
$T_{CVRP}(40, 2, 1, 0)$	0.237	100	
$T_{PC-CVRP}(40, 2, 1, 0)$	0.251	100	
$T_{CL-PC-CVRP}(40, 2, 1, 0)$	0.247	100	

Table 2.2: Fitted travel time functions.

2.4.1 Effectiveness of Systems that Pursue Proxy Objectives

Private sector planning approaches and many planning approaches proposed in the humanitarian logistics literature employ objective functions that are not necessarily in line with humanitarian objectives (Holguín-Veras et al., 2013). This might partly explain the low uptake of such approaches in the humanitarian sector. This section

addresses this hypothesis. Specifically, we compare a system that (to some extent) incorporates perceived urgency levels of requests with a system that makes planning decisions so as to minimize travel times. Travel delay in the first system is estimated using the function $T_{PC-CVRP}$ and queuing delay is estimated based on a priority queuing system (see Section 2.3). Hence, travel times are affected by priority constraints, and expected queuing delay is smaller in the high priority queue. For the travel time-based system, travel delay is estimated using the function T_{CVRP} and queuing delay is estimated based on a first-come-first-served queuing system. For this system, travel times are not hindered by precedence constraints, which comes at the expense of every request having the same queuing delay distribution.

Two important contextual determinants of the effectiveness of both systems are the travel burden and the observable variation in urgency levels. The larger the travel delay compared to the time on site, the larger the impact of travel time improvements on total delay can be. Similarly, the larger the variations in urgency levels and the better these urgency levels are assessed, the larger the impact of priority-based planning can be. Figure 2.4 depicts how relative effectiveness is affected when observable variation in urgency levels moves from low ($q_{ii} = 0.5$, $\kappa_H/\kappa_L = 1$) to high ($q_{ii} = 1.0$, $\kappa_H/\kappa_L = 1.8$) and when adapting scaling factor τ and μ^S such that, in the baseline case, the fraction of field delay spent traveling moves from 1% to 50%. All other parameters remain as described in Tables 2.1 and 2.2. The highlighted point in the figure represents the baseline case.

The results are intuitive yet interesting: (relative) effectiveness is highly context-specific, standard (private sector) objective functions can underperform in some contexts, but tend to perform comparatively well in many others. For example, our results suggest that the program under consideration would be 9.0% more effective with a travel time-based system than with a priority level-based system. This stresses the importance of carefully analyzing the context before designing a planning system.

More generally, we hypothesize that *impact-based planning*, which is to make decisions so as to minimize expected disutility, only tends to have substantial added value when both the travel burden and the need for prioritization are large. When travel times are small, routing will have little effect on field delays, so focusing on prioritization solely will yield close to optimal decisions. Similarly, when variations in urgency

levels are very small or cannot be observed, there is little *a priori* need for prioritization, and travel time-based planning will be close to optimal. When both travel times and observable variations in urgency levels are small, routing and prioritization will both have little impact on effectiveness, and a simple *heuristic planning* policy, e.g. assigning vehicles to requests in the order of requisition, will be close to optimal. We hypothesize that substantial effectiveness gains could be achieved by jointly optimizing routing and prioritization *only* when both travel delays and urgency level variations are large. Furthermore, we hypothesize that software requirements and the need to build specialized software vs. using off-the-shelf software makes impact-based planning substantially more costly than travel time-based planning and priority level-based planning, which will in turn be substantially more costly than heuristic planning. Based on these considerations, we arrive at the context-specific optimal planning approaches hypothesized in Figure 2.5.

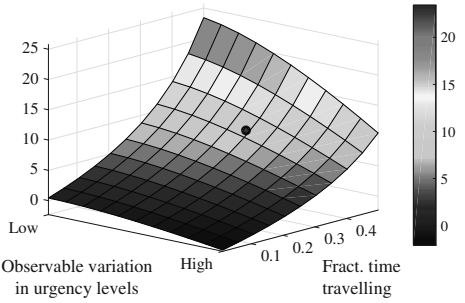


Figure 2.4: Disutility increase (%) when implementing priority-based planning instead of travel time-based planning. The highlighted point represents the baseline case.

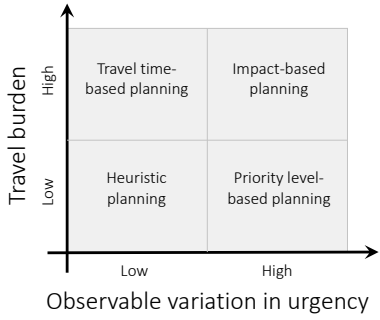


Figure 2.5: Hypothesized context-specific optimal planning criteria.

2.4.2 Effectiveness of Data-Intensive Systems

Section 2.2 stresses that information about the operational context is often lacking at a central level. Data-intensive systems might therefore be ineffective in certain

humanitarian contexts. In what follows, we investigate how the quality of travel time data and urgency assessments affects the effectiveness of data-intensive centralized planning systems. We compare a basic travel time-based system that has real-time information about travel times with a system that makes decisions based on *expected* travel times and prioritizes requests based on assessments of quality $q_{ii} = q$. The first could represent the case when local information from staff and/or drivers is incorporated during planning sessions. For a given stochasticity level δ, travel delay in this system is estimated using the function T_{CVRP} and queuing delay is estimated based on a first-come-first-served queuing system. For the data-intensive system, travel delays are estimated by the function $T_{PC-CVRP}$ with $\delta = 0$ and queuing delays are calculated based on a priority queuing system. Hence, in contrast with the basic system, the data-intensive system is able to incorporate urgency levels into planning decisions. It may, however, encounter increased travel delays due to precedence constraints and due to a gap in information about travel times.

Figure 2.6 depicts the expected increase in disutility when implementing the data-intensive system instead of the basic system, as a function of the quality of information. Specifically, we let assessment quality and operational uncertainty vary from low ($q = 0.5$, $\delta = 0.0$) to high ($q = 1.0$, $\delta = 1.0$). All other parameters remain equal.

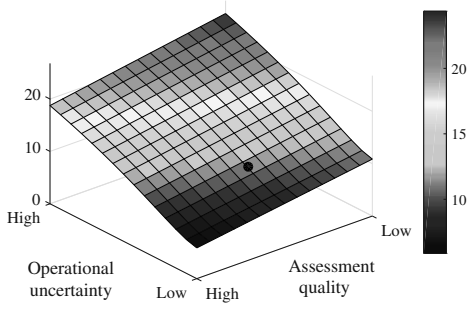


Figure 2.6: Disutility increase (%) when implementing advanced, data-intensive planning instead of basic travel time-based planning. The highlighted point represents the baseline case.

The figure shows that data-intensive systems underperform in any of the settings considered, and that the optimality gap can be rather substantial when quality of data

worsens. For example, under current levels of assessment quality and stochasticity, disutility would be 11.6% larger for the data-intensive system. The performance gaps might be smaller when there is more variation in urgency levels and when the basic system also deals with information gaps. Yet, our results point at an important general issue: centralizing decisions and incorporating more information does not necessarily lead to better decisions, particularly when data are imprecise and when local staff is already quite good at making routing decisions. Again, this stresses the importance of context-specific decision making. The next section further investigates optimal planning system hierarchies.

2.4.3 Optimal Level of Centralization

As argued, centralized systems lend themselves well to prioritization but may suffer from substantial information gaps and hence suboptimal routing. Decentralized systems are less suited to prioritization, but may have more accurate information. Hybrid systems combine the strengths of both systems by making routing decisions at the field level, where most information is available, while also incorporate prioritization. A weakness of this system is that field staff may not be able to optimally incorporate both travel time information and urgency levels. Hence, the planning problem may either need to be simplified, e.g. by taking cluster-based routing decisions, or not be solved to optimality.

Again, contextual factors largely determine the relative effectiveness of these systems. Uncertainty determines the size of the information gap centralized systems encounter, urgency levels determine the need for prioritization, and the travel burden determines the need for routing optimization. Let us assess how these factors relate to the optimal planning system. As before, we model a centralized system as a priority queuing system basing decisions on expected travel times (i.e., $T_{PC-CVRP}$ for $\delta = 0.0$). A decentralized system is represented by a first-come-first-served queuing system with travel time function T_{CVRP}. In the hybrid system, travel delays are based on the function $T_{CL-PC-CVRP}$ and queuing delays are estimated based on a priority queuing system. This system hence represents the case where the central planner divides destinations into four clusters and lists their priorities whereas

field staff optimizes routes based on actual travel times, clustering, and precedence constraints.

Figure 2.7 depicts how the optimal system changes when the information gap on a centralized level moves from low ($\delta = 0.0$) to high ($\delta = 1.0$) and when the importance of prioritization versus the importance of routing moves from low ($q_{ii} = 0.5$, $\kappa_H/\kappa_L = 1$, $\tau = \tau_{0.50}$) to high ($q_{ii} = 1.0$, $\kappa_H/\kappa_L = 25$, $\tau = \tau_{0.01}$). The three areas depicted in the figure will change, e.g., when also the decentralized and hybrid systems deal with information gaps or are incapable of optimizing routes. Yet, the figure supports the general insights drawn before: the optimal planning system is heavily context-specific and centralization does not necessarily lead to better decisions. Given that available decision support models typically assume high levels of centralization (Ortuño et al., 2013; De la Torre et al., 2012), this resonates with our second hypothesis that many of the tools proposed in the literature are not well-suited to the humanitarian context.

Our results more generally hint at the relationship hypothesized in Figure 2.8. Centralized systems outperform others when information gaps are small and when prioritization is very important. This may well represent the context of emergency medical service provisioning in developed countries, where centralized planning systems are common (Andersson and Värbrand, 2007). When uncertainty increases, hybrid systems are to be preferred, as they both incorporate adequate travel time information and account for urgency levels. This may well reflect disaster relief settings, where both uncertainty and urgency levels are high (Holguín-Veras et al., 2012). Decentralized systems maximize effectiveness when prioritization is relatively unimportant and uncertainty is large, which occurs in development assistance settings (Holguín-Veras et al., 2012). Finally, differences in effectiveness will be minor when both uncertainty and the need for prioritization is small, since each system has access to accurate information and yields near-optimal decisions by minimizing travel delays only.

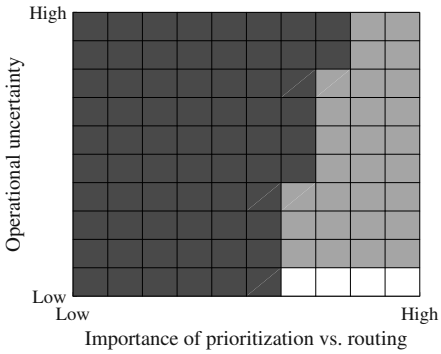


Figure 2.7: Contexts in which centralized (white), hybrid (light grey), and decentralized (dark grey) systems maximize effectiveness.

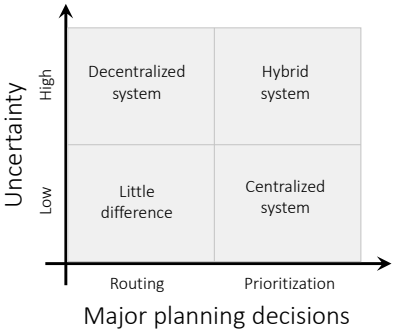


Figure 2.8: Hypothesized planning hierarchy that maximizes effectiveness.

2.4.4 Competing Planning System Adaptations

As follows from the framework depicted in Figure 2.2, optimizing planning and routing decisions is not the only way to increase the effectiveness of a planning system. Alternatives are to reduce request delays, to improve car sharing, and to increase mission lengths (i.e., the number of destinations visited per trip). For decision makers, it is relevant to know the relative potential impact of such adaptations, as this enables them to optimize their project portfolio. Assessment of these alternatives can help up assess our third hypothesis: that innovations other than optimizing planning and routing can yield greater benefit.

In what follows, we assess the impact of the different adaptations by calculating effectiveness of several systems. The first represents a simple travel time-based planning system, which we consider as a default system. The second is a system that makes *optimized planning* decisions. To obtain a rough estimate of the potential improvement, we consider the case where one can prioritize without losing routing efficiency. That is, we estimate travel times using the function T_{CVRP} and we estimate queuing delays based on a priority queuing system. In the third system, we adapt *mission lengths* by increasing the number of destinations per trip from two to three. The fourth is a *dedicated system*, which splits up the default system into two

parts having $C/2$ vehicles and arrival rates $\lambda_i/2$. In the fourth system, *request delays* have been decreased by 50%. Figure 2.9 depicts the expected delays for each of the systems. Note that queuing delays are queue-specific in the optimized scheduling case.

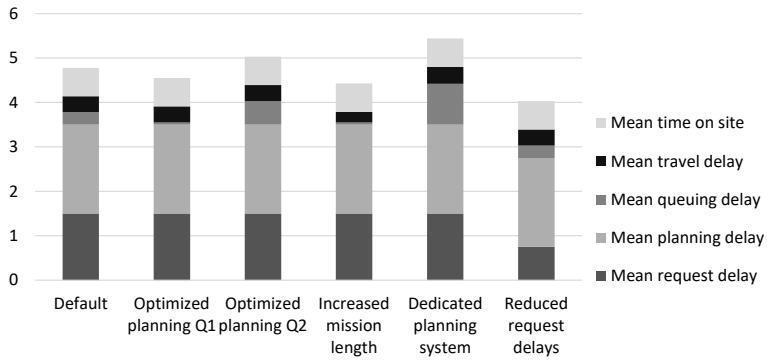


Figure 2.9: Expected delays (days) encountered between generation and fulfillment of a request for different planning systems. Q1 and Q2 denote the high and low priority queue, respectively.

The results indicate that minimizing request delays and optimizing car pooling (i.e., going from a dedicated to a pooled system) are by far the most promising strategies. Compared to the default system, effectiveness increases by 22.82% when reducing request delays and decreases by 17.04% when accounting for the assignment constraints coming with the dedicated system. Increasing mission lengths is beneficial as well, improving effectiveness by 8.28%. Optimized planning, in contrast, increases effectiveness by a mere 0.36%.

Again, one should be careful with generalizing these results to other contexts. The relative impact of different adaptations will be largely affected by contextual factors like the importance of prioritization. Yet, the order of magnitude of the effects suggests that there are a substantial number of contexts in which issues other than optimizing planning decisions deserve most attention. This also resonates with responses from several of the interviewees, who suggested that the major problems encountered on an operational level are behavioral rather than technical.

2.5 Conclusions and Discussion

Advanced planning and routing approaches are rarely employed in the humanitarian sector, despite a substantial body of literature on humanitarian last mile logistics and discontent with current practices. Major hypotheses are that (1) the effectiveness increase realized by advanced planning may not outweigh the cost of implementation, operation, and maintenance; (2) proposed approaches fail to account for essential parts of the humanitarian context; and (3) other responsibilities and innovations have a higher priority.

Our literature review, interviews with humanitarian logistics experts, and numerical analyses largely resonate with these hypotheses. The planning system that has the best balance between costs and effectiveness is highly context-specific. In some cases, determining optimal planning decisions is a trivial problem, and advanced solutions provide little added value. In other cases, advanced planning methods may have a substantial potential impact, but their cost-effectiveness is impeded by external constraints, data-quality issues, and high costs of implementation, operation, and maintenance. Centralized systems might be preferred in contexts with a high need for prioritization, but can substantially underperform when accurate information is not available, as is the case in many humanitarian contexts. For these reasons, simple decentralized planning approaches tend to be the preferred choice in many organizations. Even if installing a more advanced system is (perceived as) cost-effective, competing responsibilities tend to push this to the bottom of the fleet manager's priority list. In our case study, for example, the importance of technical issues like optimizing routing and prioritization is dominated by managerial issues such as reducing request delays and optimizing car pooling.

These issues show that it can be crucial for humanitarian organizations not to immediately join the trend in automation or decision support (cf. Vinck, 2013). There certainly is and will be room for such tools in the humanitarian sector, but research should more clearly indicate the circumstances under which this is the case. Our work should be seen as a first step into that direction. This message also applies to other decision support tools in the humanitarian sector, such as fleet optimization and fleet replacement models.

Planning approaches proposed in literature on humanitarian last-mile logistics are often weakly motivated and analyzed (see Najafi et al. (2013), Ortuño et al. (2013), and De la Torre et al. (2012) for overview articles). The literature largely fails to motivate the *complexity and data-intensity* of planning tools by consciously and pragmatically weighing costs and effectiveness. As we show, this can render their usefulness low. Complexity and data gathering may bring about substantial costs, and data-intensive systems can be very ineffective, e.g., when operational uncertainty is high, when data are imprecise or lacking, and when decisions are time-critical. Additionally, evidence showing their quality and applicability *in different humanitarian contexts* is generally lacking. Our results show that this is at least rather dangerous, as approaches that work well in one context can perform very weakly in others. Further, very few models employ evidence-based *objective functions* (cf. Holguín-Veras et al., 2013, 2012). Though our results show that systems that use proxy metrics can perform comparatively well, this does not need to be true in general. We suggest that literature in this field should do more to investigate internal and external "validity" of proposed tools. Specifically, researchers should discuss and analyze both costs and real (i.e., not proxy) effectiveness for a broad spectrum of humanitarian contexts, both for the proposed tool and for existing (common) planning practices. Only in this way can decision makers judge how much they can gain by implementing the tool in their specific context. In the same vein, we stress the importance of context-specific analyses of private sector planning approaches. Appendix 2.A argues that fundamental differences between humanitarian and private sector planning problems oftentimes exist. Yet, this does not necessarily imply that commercial planning approaches should not be used. Standard solutions tend to be cheaper than dedicated ones (Pollock et al., 2003), might work relatively well in certain humanitarian contexts (Holguín-Veras et al., 2012), and hence may be cost-effective. Research shedding more light on this matter can be very valuable.

Our model combines queuing theory with theory from vehicle routing, and provides a simple and intuitive way to analyze various planning systems. It could also serve as a basis for further modeling studies on fleet management in humanitarian organizations. Potential applications include context-specific analyses of car pooling policies, optimal vehicle allocation among different countries or programs, and opti-

mal planning periodicity. Another application would be to repeat the analyses for a variety of humanitarian organizations or programs. Classifying them within the 2x2 diagrams provides insight into the size of the quadrants and hence the relevance of different classes of planning systems and decision support tools. For our analyses to be tractable, we made several simplifying assumptions about probability distributions, their dependence, and the class of road networks, which affect our effectiveness estimates. It would be worthwhile to investigate how the model could be improved so as to obtain more precise estimates. Detailed simulation studies or studies using actual vehicle request data could be relevant future research directions. We note, however, that our findings are strongly aligned with the opinions and information obtained through our interviews and literature review and we strongly believe that our main messages remain unaffected: context is important and more advanced is not always better.

Appendices

2.A Differences with Private Sector Planning Problems

Due to economies of scale, standard planning solutions available in the private sector will likely be cheaper than solutions designed for specific humanitarian settings. From a cost-effectiveness point of view, it could hence be relevant to investigate the applicability and effectiveness of such solutions in humanitarian contexts. To assess the usefulness of private sector planning approaches in the humanitarian context, we provide an overview of several main differences between private and humanitarian sector planning problems based on our interviews and literature review. To make a specific comparison, we let the vehicle routing problem with time windows (VRPTW) represent a standard private sector planning approach. This problem is to determine routes to visit a set of destinations so as to minimize travel costs while adhering to vehicle capacity constraints and time-windows for the visits. Vehicle routing software packages typically consider variants of this problem (Baker Edward, 2002). Table 2.3

summarizes our findings. We judge a difference to be *fundamental* if it applies to objectives, constraints, decisions, the data available, and/or the nature of these data (stochastic or deterministic) and *not fundamental* if it solely applies to parameters defining a planning problem instance (e.g., the travel times, road network, number of vehicles). The reason why this distinction is made is that only fundamental differences point at issues causing the standard planning approach to potentially lose adequacy in the humanitarian context.

2.B Queuing System Analysis

M/M/C queue. The M/M/C queue represents a system with C servers, exponentially distributed service times with mean $1/\mu$, Poisson arrival of requests with rate λ, and a single queue of infinite capacity. The probability that an arriving request has to wait because all servers are busy is represented by Π and calculated as:

$$\Pi = \frac{(c\rho)^c}{c!}\left((1-\rho)\sum_{n=0}^{c-1}\frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!}\right)^{-1}. \tag{2.4}$$

Here $\rho = \frac{\lambda}{c\mu}$. Now let $\lambda^Q = c\mu(1-\rho)$. The probability density function (PDF) of the waiting time in queue X^Q, given that it is greater than 0, $f_{X^Q}(t|X^Q > 0)$, is given by (Adan and Resing, 2015):

$$f_{X^Q}(t|X^Q > 0) = \lambda^Q e^{-\lambda^Q t}. \tag{2.5}$$

M/M/C priority queue. Compared to the normal M/M/C queue, the only difference is that we now have $|J|$ priority queues. Queue j requests arrive according to a Poisson process with rates $\lambda_j, j = 1, ..., |J|$, and queue 1 requests have non-preemptive priority over queue 2 requests and so on. Again, let Π denote the probability that a randomly selected request has to wait, which is again calculated by expression (2.4) (Davis, 1966). Furthermore, let $\rho_j = \sum_{i \leq j} \lambda_i / c\mu$ (N.B. $\rho_0 = 0$).

1. Objectives: disutility of delay		
	Background:	Disutility or associated with delay of humanitarian operations typically increases convexly (Holguín-Veras et al., 2013). In typical private sector planning problems like the VRPTW, delay does not result in costs or disutility as long as service agreements on time-windows/lead times are met.
	Impact:	Given IHOs' typical aim to minimize disutility, the timing of operations is the major determinant of the quality of planning decisions.
2. Constraints: earmarked vehicle use		
	Background:	Vehicles must be used for specific purposes.
	Impact:	Total pool of requests and pool of vehicles splits up into several small pools of vehicles and corresponding requests.
3. Constraints: security issues		
	Background:	Security issues like bandits and armed conflicts form a major threat to field staff.
	Impact:	Roads can be unavailable (i.e., routes have to be changed/ the operation is to be canceled), vehicles may have to drive in convoys, and routes may need to be unpredictable.
4. Constraints: vehicle, skills, and culture requirements		
	Background:	Fulfilling a request may bring about certain requirements in terms of vehicle equipment, skills, and cultural backgrounds.
	Impact:	Assignment of vehicles and/or staff to requests may be constrained.
5. Parameters: high operational uncertainty		
	Background:	The humanitarian context typically deals with varying road and weather conditions, and demand mobility.
	Impact:	Travel times and road availability are highly stochastic.
6. Parameters: lack of operational knowledge on a central level		
	Background:	A central planner or IT system often lacks information on road conditions (affected by weather), security issues, demand mobility and their effect on driving times. Drivers and staff typically have more information.
	Impact:	Not incorporating local knowledge may induce a substantial optimality gap.
7. Parameters: stochasticity in transport availability		
	Background:	Particularly in a disaster relief context, organizations heavily rely on the local transport market. Local supply is typically highly variable.
	Impact:	Optimal planning and routing decisions may change over time with transport availability.
8. Parameters: sparse road networks, small number of destinations per mission		
	Background:	Road networks used to perform development and relief activities are typically sparse. The number of destinations visited in one mission are often small.
	Impact:	Opportunities to improve routing in these circumstances tend to be limited.

Table 2.3: Differences between private and humanitarian sector planning problems.

Davis (1966) came up with the cumulative probability function (CDF) of the waiting time of queue j, $F_{X_j^Q}(t)$. As motivated, we approximate this distribution by assuming that $X_j^Q | X_j^Q > 0$ is exponentially distributed with mean λ^{Q_j} (the real distribution is slightly different due to so-called "server vacations"). Results by Kella and Yechiali (1985) show that:

$$\lambda^{Q_j} = c\mu(1-\rho_j)(1-\rho_{j-1}), \tag{2.6}$$

and hence that the resulting conditional PDF is approximated by:

$$f_{X_j^Q}(t|X_j^Q > 0) \approx \lambda^{Q_j} e^{-\lambda^{Q_j} t}. \tag{2.7}$$

Distribution of the sum of two exponentially distributed and one conditionally exponentially distributed variable. Suppose that variables X_1 and X_2 are exponentially distributed with parameters λ_1 and λ_2 and X_3 is exponentially distributed with parameter λ_3 *given that* $X_3 > 0$. The latter occurs with probability Π. Then, if $\lambda_i \neq \lambda_j$, basic calculus yields that the PDF of the sum $X_1 + X_2 + X_3$ is given by:

$$f_{X_1+X_2+X_3}(t) = (1-\Pi)\sum_{i=1}^{2} \ell_i^{\{12\}} \lambda_i e^{-\lambda_i t} + \Pi \sum_{i=1}^{3} \ell_i^{\{123\}} \lambda_i e^{-\lambda_i t}. \tag{2.8}$$

Here, ℓ_d^D represents the Lagrange basis polynomial, as defined in (2.9).

Closed-form Expression for effectiveness. Let $\mathcal{D}_j = \{RP, Q_j, F\}$ denote the set of aggregated sources of delay a request may incur given that it ends up in queue j and let ℓ_d^D represent the Lagrange basis polynomial:

$$\ell_d^D = \prod_{d' \in D, d' \neq d} \frac{\lambda^{d'}}{\lambda^{d'} - \lambda^d} \tag{2.9}$$

Now let us consider a class i request that is assigned to queue j. Due to our assumptions, its total delay is the sum of two exponentially distributed variables, with parameters λ^{RP} resp. λ^{F}, and a conditionally exponentially distributed variable with parameter λ^{Q_j}. Let $f^j_{X_i}$ denote the PDF of the total delay distribution, as defined by (2.8). Applying basic conditioning and integration rules proves the result:

$$\begin{aligned}
\mathcal{U} &= \sum_{i\in I} \pi_i \int_0^\infty f_{X_i}(x_i) u_i(x_i) dx_i \\
&= \sum_{i\in I} \pi_i \int_0^\infty \sum_{i\in I} q_{ij} f^j_{X_i}(x_i) \gamma_i \left(e^{\kappa_i x_i} - 1\right) dx_i \\
&= \sum_{i\in I}\sum_{i\in I} \pi_{ij}\gamma_i \int_0^\infty \sum_{d\in D_j} a_{jd} \lambda^d e^{-\lambda^d x_i} \left(e^{\kappa_i x_i} - 1\right) dx_i \\
&= \sum_{i\in I}\sum_{j\in J}\sum_{d\in \mathcal{D}_j} a_{jd} \frac{\pi_{ij}\gamma_i\kappa_i}{\lambda^d - \kappa_i}.
\end{aligned}$$

where

$$\begin{aligned}
a_{jRP} &= (1-\Pi)\ell^{\{RP,F\}}_{RP} + \Pi\ell^{\{RP,F,Q_j\}}_{RP} \\
a_{jF} &= (1-\Pi)\ell^{\{RP,F\}}_{F} + \Pi\ell^{\{RP,F,Q_j\}}_{F} \\
a_{jQ_j} &= \Pi\ell^{\{RP,F,Q_j\}}_{Q_j}
\end{aligned}$$

Analysis of Queuing Delay Approximation. The following figure depicts the exact and the approximated CDF of the queuing delay distribution in the baseline case, both for the high priority queue (queue 1) and for the low priority queue (queue 2). It can be seen that the probability that an arriving request has to wait is not affected by the approximation. Also the expected queuing delay remains equal (the approximated CDF has a "fatter tail").

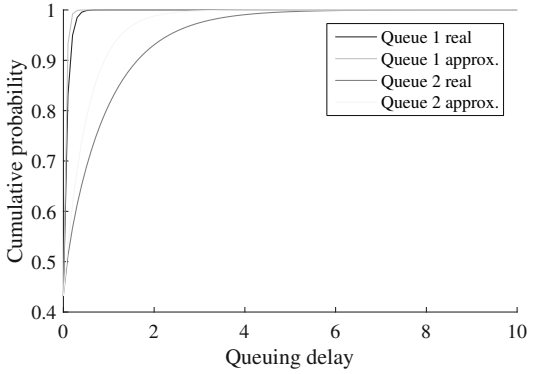


Figure 2.10: Approximated and real queuing delay distributions for the higher and lower urgency queues in the baseline case.

2.C Mean Field Delay Function

Let us consider graphs consisting of N points uniformly drawn in the square $[0, \tau]^2$. The graph includes the minimum spanning tree of a full graph plus a fraction η of the remaining edges. To mimic realistic road networks, we assume that the probability that two nodes are connected by a direct road connection decreases steeply with the Euclidian distance between them. Specifically, we assume that a given non-MST arc is included is proportional to $1/t_{ij}^3$, where t_{ij} represents the Euclidian distance. Furthermore, we assume the real travel time $\tilde{t}_{ij}$ to deviate up to a fraction δ from the expected travel time t_{ij}, according to a uniform distribution on $[t_{ij}(1-\delta), t_{ij}(1+\delta)]$.

To the best of our knowledge, literature describing the relation between expected travel times, network density, and stochasticity is lacking. Recent advances in this field are described by (Figliozzi, 2008). We therefore numerically investigate this ourselves based on two observations. First, expected travel times decrease when the network gets denser. Second, expected travel time *of a given set of routes* in the graph is independent of δ. Yet, in case that we know $\tilde{t}_{ij}$, travel times may be decreased by adapting routes. Possibilities to do so will grow for larger values of δ. These observations, combined with preliminary numerical analyses suggest that

expected travel delay per request in case of full information are well estimated by the following formula:

$$T_\varphi(N, K, \eta, \delta) \approx T_\varphi(N, K, 1, 0) \left(1 + \beta_{\varphi 1}(1-\eta)^{\beta_{\varphi 2}}\right) \left(1 - \beta_{\varphi 3}\delta^{\beta_{\varphi 4}}\right) \tag{2.10}$$

Here, $\beta_{\varphi 1}$ can be interpreted as the fractional increase in expected travel delay when routing is restricted to the MST graph instead of the full graph. Similarly, $\beta_{\varphi 3}$ represents the expected fractional decrease when having fully stochastic ($\delta = 1$) instead of deterministic ($\delta = 0$) travel times. Parameters $\beta_{\varphi 2}$ and $\beta_{\varphi 4}$ determine the "shape" of the increase and decrease. For the CVRP, $T_\varphi(N, K, 1, 0)$ can be approximated using exact estimators presented in the literature (Figliozzi, 2008; Bertsimas, 1992), after calibrating travel times (i.e. τ) to real life travel times. For the other problem variants, this can be done using simulation. Figure 2.11 illustrates the outcome obtained by least squares fitting. The depicted function yields a fit of $R^2 = 0.98$ to 121 point estimates of this function (all depicted combinations of η and δ), each of which was obtained by solving 100 randomly generated instances to optimality.

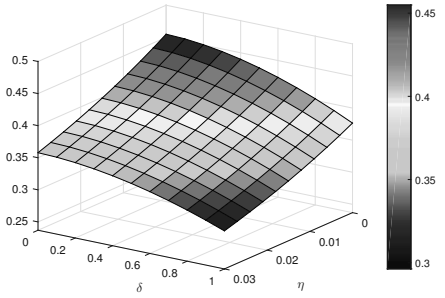


Figure 2.11: Relationship between expected travel time per request for a random CVRP instance for $N = 40$ and $K = 2$, network density (η), and stochasticity (δ).

We note that the estimated parameters will be different for different values of N and K. For example, if $K = 1$ and $\eta = 0$, the optimal solution is not affected by δ.

2.D Interviewee Characteristics

Organization	Function
CRS	Deputy Regional Director for Operations
GOAL Global	Global Fleet Advisor
	Logistics Officer
ORTEC Consulting	Vice President Global Business Consulting
	Operations Research Engineer
	Operations Research Consultant
Plan International	Logistics and Procurement Specialist
PNLTHA	Researcher
UNHCR	Chief Fleet Management Section
	Supply Chain Planning Advisor
WFP	Global Logistics Fleet Manager

Table 2.4: General interviewee characteristics. Tot ensure anonimity, the ordering of the persons and organizations is unrelated to the respondent numbers used throughout the article.

2.E Questionnaires

Base questionnaire.

Current practice	
	Who is typically in charge of planning decisions for field operations?
	Who typically makes requests for field operations?
	At what time interval are planning decisions typically made?
	How is the urgency of a request typically communicated by requestors?
	Who typically make routing decisions?
	What are the main criteria considered when making planning and routing decisions?
	What tools are being used to support routing decisions?
Effectiveness barriers	
	To what extent do the following sources of inefficiency apply to current planning and routing practices in your organization? (1) Requests for field operations/ vehicles are revealed too late. (2) Field operations are inadequately prioritized. (3) Too few destinations are combined in one mission. (4) Destinations to be combined in one mission are poorly selected. (5) Carpooling (within program or among programs) is lacking. (6) Routes taken are sub-optimal. (7) Staff does not adhere to planning and routing decisions. (8) The number of vehicles available is too low.
	To what extent do the following *operational issues* limit operational effectiveness of your organization? (1) Weather conditions. (2) Road conditions. (3) Security issues. (4) Demand mobility. (5) Donor regulations/ constraints.
	How adequate is the knowledge of field staff and drivers about operational issues?
	How adequate is the planner's knowledge about operational issues?
	How does the planner in your organization account for season-specific road conditions when making planning decisions?
	How do donor regulations/constraints affect your organization's planning and routing decisions?
	How do security issues affect your organization's planning and routing decisions?

Table 2.5: Base questionnaire. Interview questions were selected depending on the interviewee's role and organization.

Baseline case questionnaire. The interviewee was asked to consider a "typical program" within his/her organization that has its own fleet (e.g., a medium-sized program that is responsible for operations in a certain country). We also asked him/her to consider a division of all operations into two groups: the 50% of the requests having the highest urgency (i.e., the impact of delaying the fulfillment of an operation is highest), and the 50% of the requests having the lowest urgency. These two groups are referred to as "higher urgency" and "lower urgency", respectively.

How many vehicles does the typical program own/rent?
How many destinations does the program serve on average per day?
How many destinations does a mobile unit (e.g., a team in a 4x4 vehicle) visit in one mission, i.e. a sequence of visits without return to the central base? Please provide a range and an average. E.g., between 1 and 6 destinations; 2.5 destinations on average.
What is the duration of a mission? Please provide a range and an average. E.g., between 0.5 and 7 days; 1.5 days on average.
How much time does it on average take to (directly) travel from the base to a destination? Please provide a range and an average. E.g., between 30 minutes and 2 days; 4 hours on average.
How much time is there between the moment at which the need for an operation is known and the moment at which the request is revealed (e.g., to the planner/ dispatcher)? Please provide a range and an average. E.g., between 0 and 10 days; 3.5 days on average.
Suppose that we know that it takes on average 4 hours to travel between villages A and B, which are both in the area the typical program operates in. Due to varying road conditions, weather conditions, and security issues, the actual travel time from A to B may substantially deviate from this average. Please provide a range of realistic travel times. E.g., between 2 and 6 hours.
Suppose that the planner receives a request for a higher urgency operation. The planner, however, does not know the urgency level beforehand, and does an assessment based on the received information (from staff/ beneficiaries). What is the chance that a planner/ dispatcher correctly classifies the request as being of higher urgency? Please provide a percentage or a fraction. E.g., 75% or "3 out of 4 requests are correctly assessed".
Suppose you would have the choice between the following two options: (1) A higher urgency request incurs a delay of 5 days. (2) A lower urgency request incurs a delay of X days. At what value of X would you be indifferent between the two options? In other words, at what value of X would the disutility arising from delaying the lower urgency request be equal to the disutility arising from delaying the higher urgency request? E.g., at $X = 10$ days.
Which of the road network densities depicted below resembles the density of the road network the typical program operates in most closely? Here, the red circle represents the base station/ depot, the blue stars represent villages/ camps/ other destinations, and the black lines represent direct road connections between two locations.

Table 2.6: Baseline case questionnaire.

Chapter 3

The Roadside Healthcare Facility Location Problem[1]

3.1 Introduction

Sub-Saharan African truck drivers work under difficult conditions. Their trips last up to several weeks, during which time they are separated from their spouses and social-cultural norms and in which they are challenged by monotonity, risk of hijacking, loneliness, and long waiting times at border posts. This working environment has been shown to be conducive to their engagement in behaviors that bring about high risks of communicable diseases such as tuberculosis (TB), malaria, HIV, and other sexually transmitted infections (STIs) (Botão et al., 2015; Delany-Moretlwe et al., 2014; Gomez et al., 2013; Apostolopoulos and Sönmez, 2007). Prevalence levels of these diseases are reported to be high among truck drivers and their sex partners, impacting their quality of life and life expectancy (Delany-Moretlwe et al., 2014; Matovu and Ssebadduka, 2012; Apostolopoulos and Sönmez, 2007). For example, Wilson (2005) mentions a firm that lost 39 out of 144 drivers in three years. More-

[1] Apart from several minor adaptations, this chapter is a direct copy of the article: H. de Vries, J.J. van de Klundert, A.P.M. Wagelmans (2017). *The Roadside Healthcare Facility Location Problem.* Submitted to Production and Operations Management.

over, because of their mobility and their large sexual network, truck drivers have been reported to substantially contribute to the spread of HIV and other communicable diseases in Sub-Saharan Africa (Ramjee and Gouws, 2002; Gomez et al., 2013; Apostolopoulos and Sönmez, 2007; Morris and Ferguson, 2007, 2006; Laukamm-Josten et al., 2000; Caldwell et al., 1999; Hudson, 1996). A study among truck drivers in South Africa illustrates this phenomenon, showing that 56% of the 310 drivers tested were HIV positive, that 70% of them had wives or girlfriends in rural areas and that only 13% of them used a condom during their last sexual encounter. More recent studies confirm this pattern (Kiderlen et al., 2015; Botão et al., 2015; Matovu and Ssebadduka, 2012). The resulting impact on the burden and spread of these diseases translates into an economic impact. The aforementioned diseases have impoverishing effects on patients, decrease labor productivity of firms, and slow down economic growth (Sachs and Malaney, 2002; Ahlburg, 2000; Stover and Bollinger, 1999). Communicable diseases in general and HIV and other STIs in particular are also highly prevalent in truck driver populations in many other countries, e.g. in India (Pandey et al., 2008; Roa et al., 1997), Brazil (Malta et al., 2006; Lacerda et al., 1997), China (Wong et al., 2007; Chen et al., 2006), the Baltic Region (Kulis et al., 2004), and the USA (Lichtenstein et al., 2008; Solomon et al., 2004; Stratford et al., 2000).

Delivering prevention, diagnosis, and treatment services to this population for the most prevalent diseases is considered to be highly effective (Matovu and Ssebadduka, 2012; ILO, 2005; Ramjee and Gouws, 2002). Traditional healthcare facilities, however, struggle with providing the required access to truck drivers as their locations are often not easily accessible by truck, have insufficient parking space, or require detours, and since the facilities are often only open during the day whereas often drivers only have time in the evening (Delany-Moretlwe et al., 2014; Gatignon and Van Wassenhove, 2008; Ferguson and Morris, 2007; IOM, 2003; Ramjee and Gouws, 2002; Ferguson and Morris, 2007; ILO, 2005). To improve access to health services for truck drivers, NGO North Star Alliance (North Star) operates 35 Roadside Wellness Centers (RWCs) at major truck stops and border crossings along Sub-Saharan Africa's biggest transport corridors. These RWCs provide primary care services, behavior change communication, condom distribution, HIV and STI voluntary coun-

seling and testing, screening for TB and malaria, and treatment for HIV, malaria, TB, and STIs.

North Star is growing rapidly and has been employing a pragmatic expansion strategy when making the facility location decisions. Locations for new RWCs have been primarily based on patient *volume*, through estimates of the expected number of truck driver visits. Because of the long travel times and the relative sparsity of the network of roadside clinics however, the *effectiveness* of health service provisioning to truck drivers as it results from these location decisions is hampered by access discontinuities. Moreover, North Star has been offering (almost) identical health services at all of the RWCs, whereas the coverage requirements may differ per disease. As some health services are more costly (in terms of equipment, salary, and training), differentiation in service packages of the RWCs may improve the cost effectiveness of the facility location decisions. In line with common practice of leading donor organizations to target specific diseases (see e.g. Shiffman (2006) and WHO (2002)), North Star plans the introduction of a standard primary care service package, and optional packages for HIV, malaria, TB, and STI care. An optional HIV care package is already in a pilot phase (see Gomez et al., 2013). In this chapter we formulate and solve the roadside healthcare facility location problem (RFHLP): Given a limited network expansion budget, which set of locations and service packages assigned to these locations maximizes a weighted sum of effectiveness and the volume of truck drivers served?

When the only objective is to maximize patient volume, the location problem can be greedily solved to optimality by locating facilities where demand is highest. Taking effectiveness into account in the objective as well, as captured by the continuity of access measures proposed in Section 3.3, renders the defined problem to become strongly $\mathcal{NP}$-hard (as we prove in Appendix 3.E). In this chapter we set out to solve this complex problem by developing evidence-based models, and apply these models to the network of major transport corridors in South-East Africa.

Our work is the result of the network optimization project we started in 2011, together with the NGO and the ORTEC Consulting Group. The main focus of our work is on 1) evidence-based models for location decisions, 2) the development and analysis of a mixed-integer programming (MIP) formulation of the corresponding location

problem, 3) parametrizing and solving a real-life problem instance, and 4) identifying key lessons for managers and decision makers. The proposed MIP model has been implemented in a network optimization tool called POLARIS (North Star Alliance, 2011). In a subsequent project, which started in 2014, Ares et al. (2016) use our model as a basis and investigate how to incorporate equity considerations in the location problem. In contrast with our focus on *evidence-based modeling* and *management insights*, their main focus is on developing and testing a *solution method* to deal with the computational complexities induced by the equity criterion. Furthermore, their model and analysis disregards the differentiation of health services, which plays an essential role in our research, as it underlies the evidence-based modeling and our solution methods. Instead, they focus solely on the locations of clinics and assume one of the specific access measures developed in our work (CTL; see Section 3.3) to measure effectiveness. Finally, our work makes use of a substantially updated case study as it incorporates more recent demand data and adds two major corridor networks that cover eight East African countries.

The remainder of this chapter is organized as follows. Section 3.2 presents a literature review. In Section 3.3 we provide evidence-based models for location decisions, which lead to an MIP formulation of the location problem and some analytical results in Section 3.4. The results of the numerical experiments are described in Section 3.5. Finally, in Section 3.6 we summarize our findings and present conclusions.

3.2 Review of Related (Health) Facility Location Models

Our work extends the growing literature on operations research applications in global health (see Yadav (2010) and White et al. (2011) for reviews). Papers from this field typically deal with the use or allocation of scarce resources that improve health in developing countries. Kraiselburd and Yadav (2013) stress the importance of this field by describing the deficiencies in global health supply chains. The World Health Organization (WHO) also acknowledges the added value these disciplines can have for global health (Royston, 2011). Applications that use facility location models

in this context have been reviewed by Rahman and Smith (2000). In more recent work, Deo and Sohoni (2013) develop a model for allocating a health service to health clinics, Smith et al. (2009) consider the problem of locating primary healthcare workers, Griffin (2012) constructs strategies to allocate treatments to health centers and populations, and McCoy and Johnson (2014) consider the allocation of motorcycle capacity to outreach visits. McCoy and Lee (2014) investigate how to optimally allocate clinic capacity *over time*, and explicitly model the impact on HIV treatment adherence. Our problem is also related to *mobile* healthcare facility planning problems (see e.g. Doerner et al., 2007; Hachicha et al., 2000; Hodgson et al., 1998), as these problems also consider the time intervals between moments of health service provisioning. Finally, Üster and Kewcharoenwong (2011) consider the related problem of designing transportation networks so as to reduce trip lengths for the truck drivers.

The location problem described in the introduction can be regarded as a facility location problem which balances the maximization of the total node demand covered by the facilities (patient volume) with the maximization of the total flow demand covered by the facilities (which will be shown to drive effectiveness). Our problem is therefore closely related to flow interception facility location problems (FIFLPs), which have the objective of intercepting demand units on their pre-planned paths (see Boccia et al. (2009) and Hodgson (1998) for reviews of FIFLPs). Most FIFLPs assume that a demand unit is covered if and only if there exists at least one "conveniently located facility" (e.g., a facility along its path, or within a given distance from its path). The fact that it is beneficial to offer a service at multiple locations along a path makes our problem a multi-coverage FIFLP. Other than the aforementioned work of Ares et al. (2016), which builds on our work, an extensive search yielded only two other multi-coverage FIFLPs: the billboard location problem (Averbakh and Berman, 1996) and the flow refueling location problem (Kuby and Lim, 2005). Though these problems show much similarity to the RHFLP, it is not possible to directly apply models for these problems to solve our problem. As we will explain in the next sections, we measure the benefits of providing access to facilities (i.e., effectiveness) by means of a continuous variable which depends on the *travel time intervals* between adjacent facilities along a route. In contrast, the billboard location

problem measures these benefits based on the *number* of billboards along a route. The flow refueling location problem considers the driving times between adjacent facilities but differs from our problem in that it uses a binary variable to measure the benefits. For a comprehensive literature review on related facility location problems, we refer to De Vries (2011).

3.3 Continuous Access

Facilitating continuous access to healthcare along truck routes is concerned with providing levels of access that ensure effectiveness of health service provisioning. De Vries et al. (2014a) argue that the effectiveness of time-critical services such as malaria treatment and follow-up services such as treatment monitoring for HIV and TB rely on a dense network of clinics. North Star's health information system, COMETS, ensures that such network not only provides physical access to care, but also provides coordinated and integrated services at any clinic along a route.

To model the relationship between access and effectiveness, we adopt an evidence-based approach. Specifically, we follow the principles of Evidence-Based Medicine (OCEBM, 2011), which is defined as "the conscientious, explicit, and judicious use of current best evidence in making decisions" (Sackett et al., 1996). In the context of our research, these decisions do not correspond to the "care of individual patients" but the modeling of the effects of facility location and service package selection decisions, and hence fall in the domain of evidence-based (healthcare) management (see Rousseau, 2012; Walshe and Rundall, 2001). Following commonly applied evidence classification schemes, which define "a hierarchy of the likely best evidence" (Howick et al., 2011), De Vries et al. (2014a) find that an abundance of level 2, 3, and 4 evidence is available on the relationship between access and effectiveness. Appraisal of the evidence, however, reveals that the (external) validity of such evidence for the problem at hand is limited, as the corresponding study populations differ substantially from the Sub-Saharan truck driver population. Moreover, many studies find an *association* between access and health outcomes, but fail to quantify this relationship, as also recognized by Higgs (2004). Finally, existing models of access assume a fixed, static distance between a health facility and a demand unit (see Daskin and Dean

(2004), Higgs (2004) and Rahman and Smith (2000) for reviews). As the distance to facilities is constantly changing for truck drivers, the corresponding evidence lacks validity for this population. Hence, appraisal of existing best evidence makes it likely that distance importantly impacts effectiveness, but does not provide a more specific evidence base for the Sub-Saharan truck driver population. We have therefore relied on expert opinion, the 5^{th} level of evidence, to support our modelling (Section 3.3) and calibration (Section 3.5) of the relationship between access and effectiveness.

Interviews with North Star's health services experts revealed that different access requirements hold for the various diseases and corresponding service packages under consideration. We introduce a classification of service packages into three *service package types* which address these access requirements. Section 3.3.1 introduces and motivates these service package types. Next, Section 3.3.2 introduces for each of the types an access measure. Though the three package types and the corresponding access measures are largely inspired by North Star's service packages, they apply to a wide variety of health services. The mapping of North Star's service packages to the three service package types is presented in Section 3.5.1.

3.3.1 Types of Service Packages

CTL. *Service packages with a critical time-limit of access.*

This package type corresponds to diseases for which access within a time-limit from the moment of (self) diagnosis is crucial. The critical time-limit CTL type applies to rapidly progressing diseases for which timely access reduces morbidity or may even be life saving. For instance, there is considerable evidence that providing antimalarial drugs within 24 hours of onset of symptoms is essential for the treatment effectiveness (Khatib et al., 2013; Johnson et al., 2013). The dichotomous variable of having access to a health service facility within 24 hours is therefore commonly used as performance indicator (see e.g. Chuma et al., 2009; WHO/UNICEF, 2003). The CTL type may also apply to health services that need to be accessed at fixed time-intervals. One example is directly observed therapy for TB patients, in which a health provider frequently monitors a patient swallowing the required drugs (Volmink and Garner, 2007).

RCTL. *Service packages with a recommended time-limit and a critical time-limit of access.*

The CTL type implicitly assumes that one is either "too late" or "on time". For many diseases however, responsiveness is less dichotomous. The recommended & critical time-limit $RCTL$ type refers to health service packages for which a definition of being too late (i.e., a critical time-limit) and on time (i.e., a recommended time-limit) exist, but that are also characterized by a continuous relationship between accessibility and health outcomes or utilization. Strong evidence for such a *distance-decay relationship* has been confirmed in numerous studies (see e.g. McLaren et al., 2014; Cooke et al., 2010; Tanser et al., 2006). Classifying access time or treatment delay into more than two categories is very common for STIs (see e.g. Meyer-Weitz et al., 2000; Moses et al., 1994), for which there is evidence of a positive relation between treatment delay and disease progression and transmission (Hook III et al., 1997).

ASAP. *Service packages that are to be accessed as soon as possible when needed.*

The $ASAP$ type corresponds to health service packages for which there exists a distance-decay relationship without a clear time-limit of access. For instance, while there is no evidence on a specific limit for the time between diagnosis and treatment of HIV, evidence confirms that improved access is associated with less treatment delay and improved treatment adherence, which in turn lead to lower mortality and morbidity, less drug resistance, and less disease transmission (De Vries et al., 2014a; Mills et al., 2006).

3.3.2 Access Measures

We now propose access measures corresponding to the service package types introduced in the previous section. Following expert opinion and existing empirical evidence (see e.g. Ferguson and Morris, 2007; Orubuloye et al., 1993) we assume that truck drivers cyclically travel the same long distance route from which they do not deviate. Figure 3.1 illustrates the access measures for part of such a round trip of a given truck driver (i.e., a trip from origin O via destination D to origin O). The boxes (with the cross) represent a moment at which some service package s is passed.

Let us consider a given moment t during his trip. We define the truck driver's *access time to package* s as the travel time to the *next* facility along his route that offers package s. (For the reasons stated in the introduction, we assume that returning to a facility the driver already passed is not an option.) We refer to a facility that offers package s as a *package* s *location* from now on.

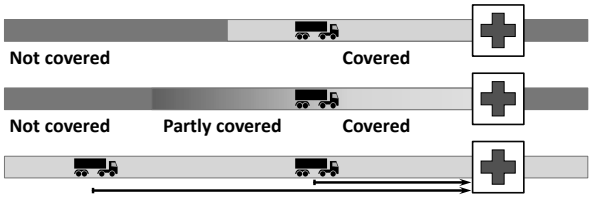


Figure 3.1: Three ways to measure continuous access. Upper: time within critical time-limit (CTL). Middle: time within recommended/critical time-limit ($RCTL$). Lower: average value of the access time ($ASAP$).

Suppose that service package s is of type CTL. At a given point in time, we say that a truck driver is "covered" if his access time to package s is at most the critical time-limit. Obviously, the larger the part of the time-line the truck driver is covered, the more effective the service provisioning can be. Therefore, we define a_s^{CTL}, the access measure for a package s that is of type CTL, as the fraction of time the truck driver is covered (see the upper part of Figure 3.1: the truck driver is covered in the light parts of the time-line). More formally, letting T denote the duration of the round trip, τ_s^{CTL} the critical time-limit for accessing package s, and $\Delta_s(t)$ the access time to package s at time $t \in [0, T]$, measure a_s^{CTL} is calculated as:

$$a_s^{CTL} = \frac{1}{T} \int_{t=0}^{T} 1_{(\Delta_s(t) \leq \tau_s^{CTL})} dt \tag{3.1}$$

Second, suppose that service package s is of type $RCTL$. Now, we say that a truck driver is "covered" if his access time is smaller than the recommended time-limit, that he is "not covered" if his access time exceeds the critical time-limit, and that he is "partially covered" otherwise. We assume that the degree of coverage with package s decreases linearly from 1 (covered) to 0 (not covered) as the access time

increases from the recommended time-limit to the critical time-limit (see the middle part of Figure 3.1: the lower the truck driver's degree of coverage, the darker the shade of grey). Next, we define a_s^{RCTL}, the access measure for a package s that is of type $RCTL$, as the average degree of coverage with package s during a trip. Formally, let the recommended time-limit and the critical time-limit of access be represented by τ_{1s}^{RCTL} and τ_{2s}^{RCTL}, respectively. Then a_s^{RCTL} is calculated as:

$$a_s^{RCTL} = \frac{1}{T} \int_{t=0}^{T} 1_{(\Delta_s(t) < \tau_{1s}^{RCTL})} + 1_{(\tau_{1s}^{RCTL} \leq \Delta_s(t) \leq \tau_{2s}^{RCTL})} \cdot \left(\frac{\tau_{2s}^{RCTL} - \Delta_s(t)}{\tau_{2s}^{RCTL} - \tau_{1s}^{RCTL}} \right) dt \tag{3.2}$$

For later use, we refer to the value of the integral $\int(\dots)dt$ in equation (3.2), taken from t_1 to t_2, as the "effective" time the truck driver is covered during the period $[t_1, t_2]$.

Finally, suppose that the service package is of type $ASAP$. Then, for each point of time during the truck driver's trip, the following holds: the lower the access time, the more effective the health service accessed can be. Based on this observation, we define a_s^{ASAP}, the access measure for a package s of type $ASAP$, as the average value of the truck driver's access time during his trip (i.e., the average length of the arrow in the lower part of Figure 3.1):

$$a_s^{ASAP} = \frac{1}{T} \int_{t=0}^{T} \Delta_s(t) dt \tag{3.3}$$

One may interpret a_s^{ASAP} as the expected travel time to the next package s location from the moment the truck driver needs this service package.

3.4 Model

In this section, we formally model the RHFLP. Some notations are introduced in Section 3.4.1. Section 3.4.2 models the value of a set of location decisions in terms of

volume of truck drivers served and effectiveness of healthcare provision. Next, Section 3.4.3 introduces a mixed-integer programming (MIP) formulation of the problem. Finally, Section 3.4.4 provides a worst case analysis with respect to data inaccuracy.

3.4.1 Notation

We model a problem instance by means of a graph $G(L, E)$, where L and E denote a set of vertices and a set of edges connecting these vertices, respectively. Table 3.1 introduces the notations we use to denote the graph, the service packages, the demand parameters, and the decision variables.

Here, a truck flow q represents a collection of truck drivers who travel the same long distance route. A route defines a path in the graph, starting from origin O_q, visiting the ordered set of facility locations K_q, and ending at destination D_q. that, although O_q (or D_q) may physically represent the same location as a facility location $k \in K_q$, we regard them as separate locations. K_q may include potential locations from the set KP as well as locations from the set of locations which are currently already in use, KC. The latter may include facilities operated by other organizations. For instance, if a facility is accessible, and contributes to continuity of access, it can be viewed as being part of the NGO's network. Finally, we note that allocation of new service packages can be restricted to the newly located facilities only ($KP_s = KP$) or to any existing facility ($KP_s = \{k : y_{ks} = 0\}$).

General:	
p	number of roadside health facilities to be newly located
p_s	number of facilities service package s can be newly allocated to
Ω	set of all feasible solutions $\omega = (\mathbf{x}, \mathbf{y}) = (\{x_k\}, \{y_{ks}\})$
Network:	
KC	set of current facility locations
KP	set of potential facility locations
KP_s	set of potential locations for package s
K	set of facility locations: $K = KC \cup KP$
O	set of truck route origins, $O \cap K = \emptyset$
D	set of truck route destinations, $D \cap K = \emptyset$
L	set of locations: $L = KC \cup KP \cup O \cup D$
Q	set of long distance truck flows
O_q	origin of flow $q \in Q$, $O_q \in O$
D_q	origin of flow $q \in Q$, $D_q \in D$
K_q	ordered set of facility locations along the route of flow $q \in Q$, $K_q \subseteq K$
χ_{ks}	parameter indicating whether a *current* facility at location k offers package s ($\chi_{ks} = 1$) or not ($\chi_{ks} = 0$)
Service packages:	
S	set of health service packages
J	set of service package types $J = \{CTL, RCTL, ASAP\}$
S_j	set of service packages that are of service package type j, $S_j \subseteq S$
e_{qs}	effectiveness of health service provisioning per truck driver in flow q demanding service package s, given solution ω, $e_{qs} \in [0, w_s]$
$a^j_{qs}(\mathbf{y})$	level of access to package s provided for truck drivers in flow q, given package allocation decisions $\mathbf{y}$ and the access measure corresponding to the selected service package type j
$\underline{\alpha}^j_s, \overline{\alpha}^j_s$	break points of piecewise linear function $e_{qs} = g^j_s\left(a^j_{qs}(\mathbf{y})\right)$
w_s	maximal attainable effectiveness per truck driver demanding package s
Demand parameters:	
d_k	expected daily patient volume at facility location k
f_{qs}	number of truck drivers in flow q who need service package s
Decision variables:	
x_k	$\begin{cases} 1 & \text{if a health facility is placed at location } k \\ 0 & \text{otherwise} \end{cases}$
y_{ks}	$\begin{cases} 1 & \text{if service package } s \text{ is offered at location } k \\ 0 & \text{otherwise} \end{cases}$

Table 3.1: Table of Notations.

3.4.2 Optimization Criteria

A solution $\omega = (\mathbf{x}, \mathbf{y}) \in \Omega$ specifies a network of roadside healthcare facilities, including all current and new facilities and their service packages. The network expansion budget allows for placing p new facilities and assigning service package s to p_s facilities. For a given solution ω, we define the patient volume Z_{PV} as the expected daily number of truck driver visits captured by the roadside healthcare facilities in this network. Hence, letting d_k denote the patient volume at location k, we derive that:

$$Z_{PV} = \sum_{k \in K} d_k x_k \tag{3.4}$$

This definition of Z_{PV} implicitly assumes that the patient volume at a given location is independent of our location decisions (i.e., by the service packages offered at this location, and by the locations of other facilities). In Section 3.5.4 we will show that impreciseness potentially induced by this assumption hardly affects the optimal location decisions.

Let us now quantify for a given solution $\omega \in \Omega$ the effectiveness of health service provisioning. We denote the number of truck drivers in flow q demanding service package s by f_{qs}. For each of these truck drivers, e_{qs} represents the effectiveness of health service package s provisioning, measured in terms of quality-adjusted life years (QALYs) or some other measure (we come back to this later). Then the total effectiveness of package s provisioning is defined as:

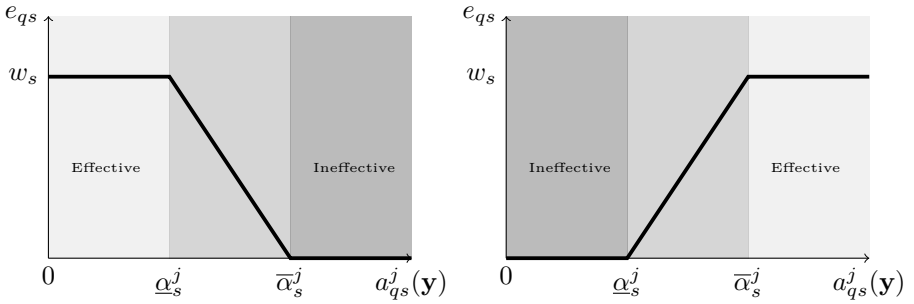
$$Z_s = \sum_{q \in Q} f_{qs} e_{qs} \tag{3.5}$$

We let functions g_s^j define the relationship between effectiveness and access. More precisely, for package s, we firstly select the most fitting of the three proposed service package types and corresponding access measures. Let $j \in \{CTL, RCTL, ASAP\}$ denote the selected service package type. Furthermore, given package allocation

decisions $\mathbf{y}$, let $a^j_{qs}(\mathbf{y})$ denote the level of access to package s for flow q, as measured by the access measure corresponding to type j (see Section 3.3.2). Then, the effectiveness of package s provisioning for truck drivers in flow q is measured as:

$$e_{qs} = g^j_s\left(a^j_{qs}(\mathbf{y})\right) \qquad s \in S_j \tag{3.6}$$

These functions $g^j_s(\cdot)$ reflect which levels of access are ineffective, partially effective, and effective. Lacking empirically or theoretically validated alternatives and for ease of modeling, we define $g^j_s(\cdot)$ as piecewise linear functions. Figures 3.2a and 3.2b illustrate these functions. In case that $a^j_{qs}(\mathbf{y})$ is to be maximized (i.e., when $j \in \{CTL, RCTL\}$), we define that e_{qs} *increases* linearly from 0 to the maximal attainable effectiveness w_s when $a^j_{qs}(\mathbf{y})$ increases from a given lower bound threshold $\underline{\alpha}^j_s$ (ineffective access) to a given upper bound threshold $\overline{\alpha}^j_s$ (effective access). In case that this variable is to be minimized (i.e., when $j = ASAP$), e_{qs} *decreases* linearly from w_s to 0 when $a^j_{qs}(\mathbf{y})$ increases from the lower bound threshold $\underline{\alpha}^j_s$ (effective access) to the upper bound threshold $\overline{\alpha}^j_s$ (ineffective access). Section 3.5.1 discusses the calibration of these parameters.



(a) Function for $j = ASAP$. **(b)** Function for $j \in \{CTL, RCTL\}$.

Figure 3.2: Effectiveness of package s provisioning as a function of $a^j_{qs}(\mathbf{y})$.

Parameters w_s can represent the health value of providing one truck driver demanding service package s with effective access to this package, which could be

expressed in terms of quality-adjusted life years (QALYs). In that case, e_{qs} represents a health value estimate in QALYs as well. Alternatively, the values can be chosen so as to mimic the strategic objectives of the NGO and/or its donors, specifying the disease(s) to focus the investments on. In the latter case, effectiveness is to be interpreted as the extent to which a solution meets these objectives. We note that, under the assumption that decision makers can determine the "optimal" balance between the two main optimization criteria, multiplying w_s by a given constant does not change optimal location decisions. Consequently, it suffices if w_s represents the *relative* maximal attainable effectiveness of package s provisioning, which is significantly easier to estimate than absolute effectiveness.

For later use, we represent the total effectiveness of health service provisioning by Z_E, and refer to this variable as the *effectiveness score*:

$$Z_E = \sum_{s \in S} Z_s \tag{3.7}$$

3.4.3 MIP Formulation

The measures presented in Section 3.4.2 allow us to introduce the following problem formulation, which we refer to as the *Roadside Healthcare Facility Location Problem* (RHFLP). Here, parameters $r \in [0, 1]$ and $(1 - r) \in [0, 1]$ capture the relative importance of the patient volume criterion and the effectiveness criterion, respectively:

$$\max \quad r \sum_{k \in K} d_k x_k + (1-r) \sum_{s \in S} \sum_{q \in Q} f_{qs} e_{qs} \tag{3.8}$$

$$\text{s.t. } e_{qs} = g_s^j \left(a_{qs}^j(\mathbf{y}) \right) \qquad q \in Q, j \in J, s \in S_j \tag{3.9}$$

$$\sum_{k \in KP} x_k = p \tag{3.10}$$

$$\sum_{k \in KP_s} y_{ks} = p_s \qquad s \in S \tag{3.11}$$

$$x_k \geq y_{ks} \qquad k \in K, s \in S \tag{3.12}$$

$$x_k = 1 \qquad k \in KC \tag{3.13}$$

$$y_{ks} = \chi_{ks} \qquad k \in KC, s \in S \tag{3.14}$$

$$x_k, y_{ks} \in \{0,1\} \qquad k \in K, s \in S \tag{3.15}$$

The objective function (3.8) maximizes a weighted sum of Z_{PV} and Z_E. Constraints (3.10) and (3.11) specify the number of new facilities and the number of service packages s to be allocated to potential facility locations. Next, constraints (3.12) define that a service package can only be allocated to facility location k if a health facility is located there. The current network of health facilities is described in constraints (3.13) and (3.14). Constraints (3.15) define our decision variables as binary variables. Last, the effectiveness variables e_{qs} are defined in equations (3.9).

The non-linearity of these equations makes (3.8) - (3.15) a mixed-integer *non-linear* programming (MINLP) model, which is generally very hard to solve. In Appendices 3.A - 3.D, however, we show how a series of modeling tricks transforms this model into a mixed-integer *linear* programming (MILP) model. Specifically, it suffices to replace equations (3.9) by constraints (3.17) - (3.20) and (3.22) - (3.36). We also prove that the vast majority of the binary variables needed to perform this transformation can be relaxed to continuous variables, since the corresponding constraint matrix is totally unimodular (TU). This thereby substantially decreases the complexity of solving the MILP model.

Finally, in Appendix 3.E, we prove the following by reduction from the problem CLIQUE:

Proposition 3.1. *The RHFLP, as formulated by* (3.8) - (3.15)*, is strongly $\mathcal{NP}$-hard, even in case that $|S| = 1$.*

3.4.4 Worst-case Effect of Imprecise Input Data

Parameters f_{qs}, the number of truck drivers in flow q who need package s, and d_k, the expected daily patient volume at location k tend to be imprecise in the data-scarce environment North Star operates in. Though this does not affect the feasible solution space, it may affect the value of a solution and thereby induce an optimality gap.

Proposition 3.2 provides an upper bound on this gap (see Appendix 3.F for the proof). Here, $\omega^{\#}$ denotes the true optimal solution and ω^* the optimal solution based on the presently used parameter values. We denote the value function using the true parameters by $v^{\#}$, and the value function using the presently used parameters by v^*. For instance, $v^{\#}(\omega^*)$ denotes the value of the solution ω^* in case that the true parameter values were used.

Proposition 3.2. *If the true values of f_{qs} and d_k deviate at most by a fraction δ from the presently used parameter values, then $\frac{v^{\#}(\omega^{\#})}{v^{\#}(\omega^*)} \leq \frac{1+\delta}{1-\delta}$. This bound is tight.*

To illustrate this bound, let δ be equal to 0.1. Then in the worst case, the true value of ω^* is about 18% lower than the true value of $\omega^{\#}$. As this bound is only attained with equality under three very specific circumstances (see Appendix 3.F), Section 3.5.4 investigates realistic levels of sensitivity by means of a real life case study.

Let us now analyze the optimal solution's sensitivity with respect to the parameters w_s, the maximal attainable effectiveness of package s provisioning, and parameters r and $1 - r$, the relative values of the patient volume criterion and the effectiveness criterion, respectively. Suppose that there exists an "optimal" value of these parameters. For example, optimal values could be those values yielding the largest health improvement. As it is very difficult to identify these values, it is likely

that the chosen parameter values deviate from them. An upper bound on the resulting optimality gap is given in Proposition 3.3 (see Appendix 3.G for the proof). Here, $w_{ob} \in \{r\} \cup \{(1-r)w_s\}$ and γ denotes some instance-dependent constant between 0 and 1.

Proposition 3.3. *If the "optimal" values of w_{ob} deviate at most a fraction δ from the presently used parameter values, then $\frac{v^{\#}(\omega^{\#})}{v^{\#}(\omega^*)} \leq \frac{(1+\delta)^2}{(1-\delta)\cdot(1+\delta+2\delta\gamma)}$. This bound is tight.*

For example, if $\gamma = 0.5$ and $\delta = 0.1$, then in the worst case, the optimality gap is at most 11%.

In computational experiments presented in the next section, we compare these bounds with the actual bound *for a given instance*. It is easy to see that, in the worst case, the true value of a given parameter is either the fraction δ larger or the fraction δ smaller than the presently used values. Let an extreme realization of the parameters of interest be represented by $\psi \in \Psi$, and let $v^{\psi}(\omega)$ denote the resulting value function. Then the worst case optimality gap for a given instance can be found by solving the following problem:

$$\max_{\omega \in \Omega, \psi \in \Psi} v^{\psi}(\omega) - v^{\psi}(\omega^*) \tag{3.16}$$

3.5 Impact Analysis

This section illustrates how our model can be used for strategic planning and reveals practical insights for decision makers. We use POLARIS, the software package we built for North Star (North Star Alliance, 2011), to solve our MILP model. This package uses CPLEX 12.6.2 as a solver and was run on a PC with a 3.4 AMD A4-5300 processor and 8 GB RAM. Section 3.5.1 describes our baseline case study and Section 3.5.2 investigates several network expansion strategies. As proven, the RHFLP is a strongly $\mathcal{NP}$-hard problem, even in case that $|S| = 1$. In Section 3.5.3, we apply our model to a variety of cases to get more insight into the tractability of the model. Next, in Section 3.5.4 we numerically investigate the sensitivity of the

optimal solution with respect to impreciseness in the demand data, and compare the results with the bounds obtained in Section 3.4.4. Finally, Section 3.5.5 explores how changes in the objective weights and in the number of new facilities to locate affect the optimal solution.

3.5.1 South-East Africa Case

We base our test case on the network of major transport corridors in South-East Africa. This network of roads spans 14 Sub-Saharan countries and connects the main ports in the region with the main inland cities and other areas of economic importance (e.g., the copper belt in the DRC). It consists of three subsets of corridors: (1) the North-South Corridor, which spans the region from Tanzania to South Africa, (2) the Central Corridor, which runs from the port of Dar es Salaam (Tanzania) into Burundi, Rwanda, and the DRC, and (3) the Northern Corridor, which connects the port of Mombasa (Kenya) with landlocked countries lying east and north of Kenya. Figure 3.3 depicts the corridor network as a graph consisting of the 29 RWCs North Star currently operates in this network and 85 potential RWC locations.

The case study data sources are summarized in Table 3.2. In line with package allocation choices currently faced by North Star, we choose the primary care (PC) package to be a standard service package and the HIV care (HC) package to be the only optional service package. Hence, $S = \{PC, HC\}$. The data sources identify 95 major truck flows, and their health service needs are calculated as $f_{qs} = f_q \theta_{qs}$. Here, f_q and θ_{qs} denote the number of truck drivers in flow q and the fraction of truck drivers in flow q who need service package s, respectively. For truck flows making use of the Northern and the Central Corridor, estimates of f_q were obtained from inter-country cargo volumes presented by Nathan Associates Inc. (2011) and Berger (2011), respectively. For the North-South Corridor, we use least squares optimization to fit f_q to annual average daily truck volume (AADT) estimates ($R^2 = 0.93$), which Odoki and Anyala (2014) provide for most arcs in the network, and to a "target matrix" describing road-based inter-country cargo volumes ($R^2 = 0.81$), as presented by the SADC (2012). The use of such a target matrix is a common way to deal with the limitations of OD matrix estimation methods (Peterson, 2007). We choose $\theta_{q,PC} = 1$,

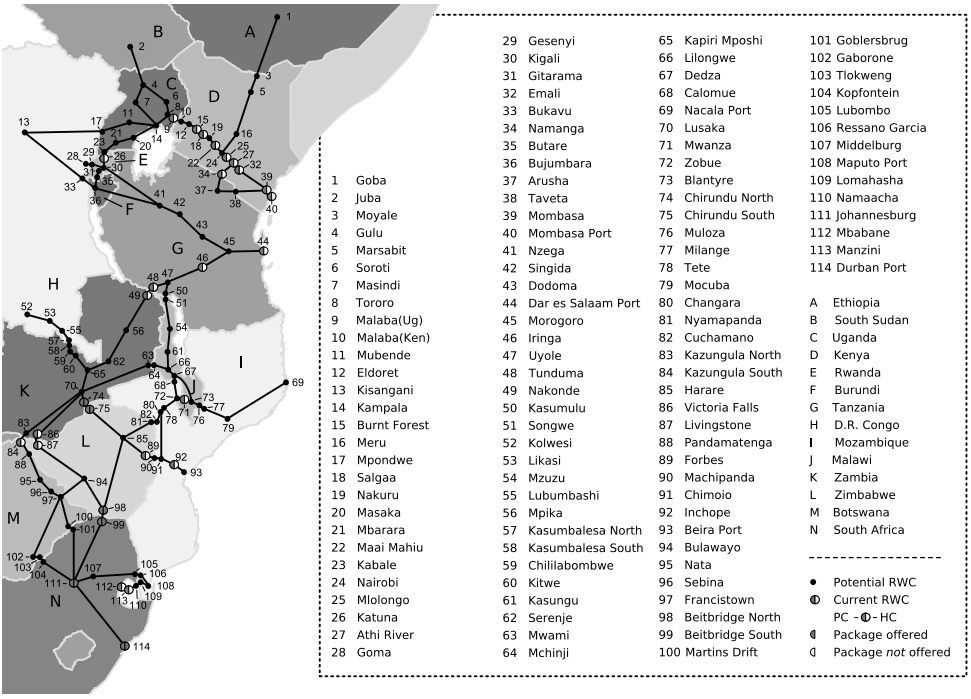


Figure 3.3: Map of the South-East African Transport Corridor Network.

reflecting the assumption that every truck driver has a potential need for primary care, and estimate $\theta_{q,HC}$ as 1.41 times the average of the adult HIV prevalence in the flow's origin country and the prevalence in the flow's destination country. The multiplier 1.41 reflects the difference between the prevalence levels in the general population and the truck driver population, and is based on data for South Africa and Mozambique. A truck flow is assumed to take the shortest route from its origin to its destination.

Finally, we use least squares optimization to fit the relationship between truck volume along current RWC locations ($AADT_k$) and daily patient volume at these locations (d_k) ($R^2 = 0.33$). This yields the following estimation method for patient volume at *potential* RWC locations: $d_k = 10.65 + 0.31 \cdot \sqrt{AADT_k}$.

Data	Source	Value
Network:		
KC, χ_{ks}	North Star, Gomez et al. (2013)	
KP	Inspection of corridor network	
$Q, t_{kl}, AADT_k$	North Star (via partner organizations), TTCA (2015), Odoki and Anyala (2014); JICA (2013, 2010), SADC (2012); Ranganathan and Foster (2011), Nathan Associates Inc. (2011); Berger (2011) Teravaninthorn and Raballand (2009)	
Service packages:		
S	North Star	Table 3.3
S_j	Expert survey	Table 3.3
w_s	Expert survey	Table 3.3
τ_{1s}^{RCTL}, τ_{2s}^{RCTL}	Expert survey	Table 3.3
$\underline{\alpha}_s^j$, $\overline{\alpha}_s^j$	Expert survey	Table 3.3
Demand parameters:		
d_k	North Star	
f_q	Odoki and Anyala (2014); SADC (2012) Nathan Associates Inc. (2011); Berger (2011)	
$\theta_{q,PC}$	Assumption	
$\theta_{q,HC}$	WHO (2016); Botão et al. (2015) Delany-Moretlwe et al. (2014)	

Table 3.2: Data sources baseline case.

As mentioned in Section 3.3, our calibration of the relationship between access and effectiveness is based on structured expert surveys. Specifically, we asked two experts from North Star to quantify the relative importance of the two service packages (yielding w_s), to connect to each service package a corresponding service type (yielding S_j), and to assign scores to 10 configurations of RWCs offering the PC package along an imaginary truck route and to 10 configurations of RWCs offering the HC package along an imaginary truck route. The functions $g_s^j(\cdot)$ providing the best fit to these scores are determined by least squares optimization ($R^2 = 0.77$ for PC, $R^2 = 0.69$ for HC). The resulting parameter values are given in Table 3.3.

Under the assumption that decision makers can identify the "optimal" balance between the two main criteria, as quantified by the parameter r, it suffices that parameters w_s represent the *relative* importance of the different packages (i.e., the scaling of these parameters does not matter). Since experience suggests that decision makers will vary the parameter r in the software package we developed (see

North Star Alliance, 2011) until one or more satisfactory solutions are obtained, we believe this assumption to be reasonable.

The service type allocation obtained from the expert surveys – PC to the $RCTL$ type and HC to the $ASAP$ type – is largely influenced by the time criticality of the corresponding health services, the aforementioned distance decay relationship, and the benefits of repeatedly passing a health facility:

> **Expert 1 about truck drivers accessing the PC package (translated):** *There is a wide variety in conditions (...). Generally, the people do not go to the clinic for nothing. Certainly not the truck drivers. They have a problem that builds up till a certain critical level. Then they decide to come, and if they cannot access a clinic within two days, they have a very big problem.*
>
> **Expert 2 about the benefits of continuous access to the HC package (translated):** *One is biomedical, two is psychological – a sense of urgency among the patients – (...) and there is a third (...) the power of repetition is in that, if he thinks "should I do it or not" [go to a clinic] (...) the third time [he passes a clinic] he thinks "oh no, now I'm gonna do it" (...). So when it comes to HIV, you never have too many [clinics].*
>
> **Expert 1 about the time criticality of the HC package (translated):** *The problem evolves rather gradually. Problems arise when they run out of medication. Then you got a problem... then you need to come (...). When they get an opportunistic infection (...) it builds up rather slowly. (...) Also in The Netherlands, when you have this problem [HIV infection], you don't see a specialist within two days.*

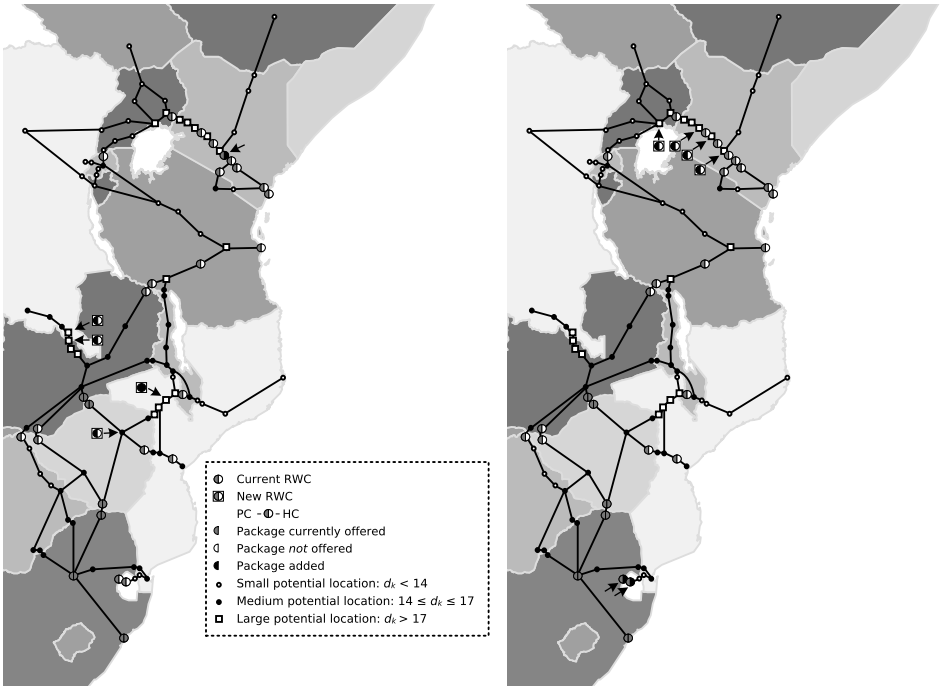
Package s	Service package type j	Access metric parameters (days)		$\underline{\alpha}_s^j$	$\overline{\alpha}_s^j$	w_s
PC	$RCTL$	$\tau_{1PC}^{RCTL} = 1.29$	$\tau_{2PC}^{RCTL} = 3.00$	0.09	1.00	$1 \cdot \sigma$
HC	$ASAP$			0.40	5.48	$2 \cdot \sigma$
MC	CTL	$\tau_{MC}^{CTL} = 1.00$		0.50	0.90	$5 \cdot \sigma$
TC	$RCTL$	$\tau_{1TC}^{RCTL} = 1.00$	$\tau_{2TC}^{RCTL} = 2.00$	0.50	0.90	$2 \cdot \sigma$
SC	$RCTL$	$\tau_{1SC}^{RCTL} = 1.00$	$\tau_{2SC}^{RCTL} = 2.00$	0.50	0.90	$1 \cdot \sigma$

Note: the TC, SC, and MC packages are only used in Section 3.5.3. W.l.o.g. we set scaling factor $\sigma = 1/25$

Table 3.3: Parameters describing the relationship between access and effectiveness of package s provisioning.

3.5.2 Network Expansion Strategies

In our baseline case, we optimize locations of 4 new RWCs and the allocation of 2 HC packages among any established RWC (i.e., $p = p_{PC} = 4$, $p_{HC} = 2$). Based on this case, we analyze two network expansion strategies, characterized by the parameter r. The first strategy is the "balanced strategy", which we model by setting $r = 0.5$ (many Pareto efficient solutions that balance both objectives are found around this value, as we show in Section 3.5.5). The second investment strategy mimics the strategy North Star had been employing before this research was implemented. The main focus of this strategy was put on maximizing the patient volume served (the effects in terms of enhancing continuity of access were not explicitly considered), which we model by setting $r = 1.0$. We refer to this strategy as the "patient volume strategy". Finally, we also consider the case in which we not only locate new RWCs and service packages, but also allow current clinics to be relocated. Here, we again use $r = 0.5$.



(a) Balanced solution. **(b)** Patient volume solution.

Figure 3.4: Optimal location decisions for the baseline case.

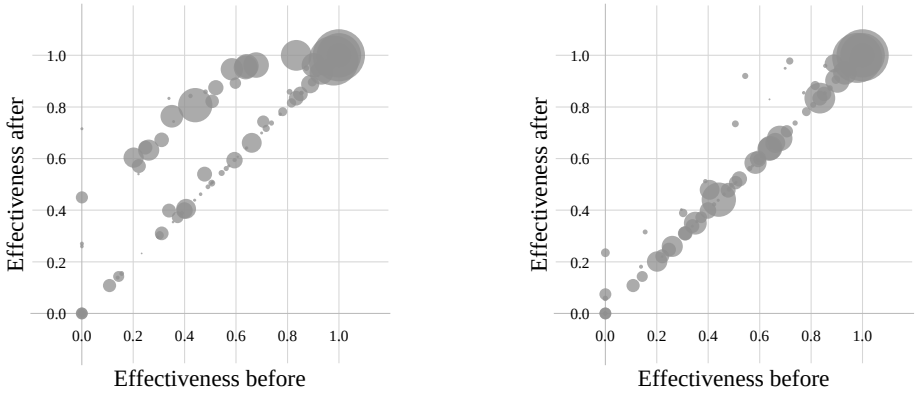
Solution	Z_{PV}		Z_{PC}		Z_{HC}	
Initial	562		19840		1646	
Balanced	632	(12.3%)	22964	(15.7%)	2353	(43.0%)
Patient volume	662	(17.6%)	20070	(1.2%)	1652	(0.4%)
Redesigned	682	(21.2%)	24342	(22.7%)	2673	(62.4%)

Note: The percentage improvement with respect to the initial network is given between brackets.

Table 3.4: Daily patient volume (Z_{PV}), effectiveness of primary care provisioning (Z_{PC}), and effectiveness of HIV care provisioning (Z_{HC}) in the initial network, the networks obtained by the balanced strategy and the patient volume strategy, and in the redesigned network.

Figures 3.4a and 3.4b and Table 3.4 describe the optimal solutions yielded by the balanced strategy and the patient volume strategy. We see that the "balanced solution" results in a relatively lower expected daily patient volume (5.3 percentage

points lower) but also in a substantially higher effectiveness (14.5 and 42.6 percentage points higher for PC and HC, respectively). The latter is illustrated by the bubble plots in Figures 3.5a and 3.5b, which depict the solutions' impact on the effectiveness of package PC provisioning for the 95 truck flows separately. The reason is that the patient volume solution places RWCs in the busiest parts of the network, which are already well-served by current RWCs (see Figure 3.4b). The balanced strategy, instead, yields locations that fill up gaps in the RWC network. For example, Figure 3.6 shows how the new RWCs improve the degrees of coverage with the PC package during trips along the routes connecting the ports of Durban, Beira, Nacala, and Dar es Salaam with the copper belt (Kolwesi, DRC).



(a) Balanced solution. **(b)** Patient volume solution.

Figure 3.5: Effectiveness of package PC provisioning for the 95 truck flows before and after implementing the balanced solution and the patient volume solution. A bubble's size reflects the number of truck drivers in the corresponding flow.

The optimal network obtained when relocation of clinics is allowed moves 17 of the current 29 clinics and all of the 6 current HC packages to a different location. Compared to the case when relocation is not allowed, this significantly increases patient volume (8.9 percentage point increase) and the effectiveness of PC provisioning (7.0 percentage point increase) and HC provisioning (19.4 percentage point

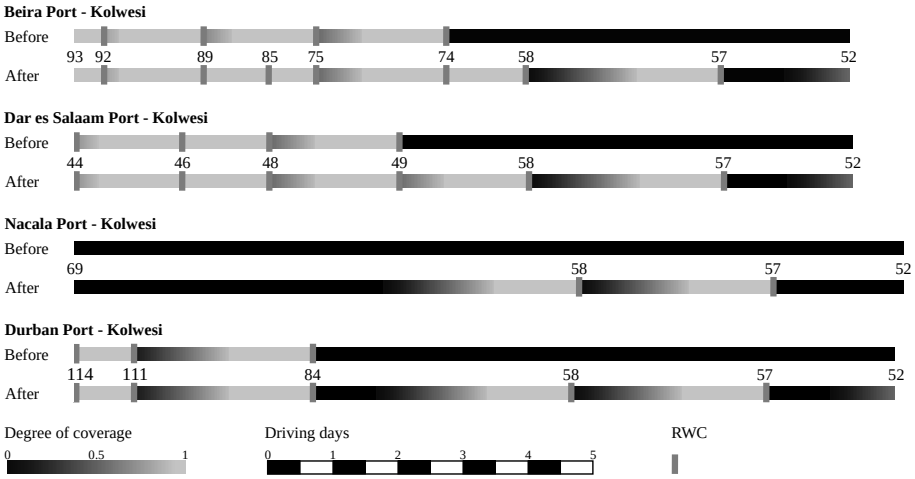


Figure 3.6: Degree of coverage with the *PC* package before and after locating the new RWCs at each point of time during a trip from a major regional ports (Durban, Beira, Nacala, and Dar es Salaam) to the copper belt (Kolwesi, DRC). The numbers correspond to the location numbers introduced in Figure 3.3.

increase). This provides a strong argument for critically assessing the fitness of the current network.

Finally, the optimal solutions illustrate an important property of the RHFLP: adding multiple packages to the network can bring about *synergy effects*. For instance, when adding the *PC* package to Kasumbalesa North or to Kasumbalesa South only, the effectiveness of package *PC* provisioning to truck flow Nacala Port - Kolwesi increases by 0.09, whereas it increases by 0.27 if this package is added to *both* locations. This is a direct implication of the assumption that a certain level of access is required before this translates into a strictly positive level of effectiveness. These effects have important implications for decision makers, which will be discussed in Section 3.6.

3.5.3 Model Statistics

Table 3.5 describes the solution times and model statistics of 12 problem instances. The problem instances are defined by the number of truck flows (55 or 95), the number of service packages (2 or 5), and the structure of the network (baseline, dense, or sparse). The three additional packages are future optional packages: tuberculosis care (TC), STI care (SC), and malaria care (MC). Table 3.3 shows the definitions of the corresponding effectiveness variables, which were based on expert discussions (service package type allocations) and assumption (other parameters). The fractions of truck drivers who need these packages were again estimated based on data from the global health data repository (WHO, 2016). The sparse network is generated by selecting the minimum spanning tree of our graph $G(L, E)$, and the dense network is generated by adding for each vertex two imaginary edges that connect this vertex with the two closest neighbor vertices it was not connected to in the baseline network. We generate the time needed to traverse such imaginary edge by dividing the Euclidian distance of the edge by the average speed of 40 km/hour. For both alternative network structures, we determine the shortest origin-destination routes. All other parameters are kept as before.

Instance	$\|Q\|$	$\|S\|$	Structure	Variables (Int.)	Constraints	LP gap (%)	CPU (sec.)
D55	55	2	Dense	5679 (558)	2933	0.47	0.6
D95	95	2	Dense	9123 (798)	4825	0.50	2.1
B55	55	2	Baseline	9769 (558)	3625	2.15	26.5
B95	95	2	Baseline	15871 (798)	5977	1.85	4.6
S55	55	2	Sparse	13725 (558)	4109	3.01	88.5
S95	95	2	Sparse	22665 (798)	6837	3.33	136.7
D55e	55	5	Dense	14196 (1395)	7502	3.35	19.6
D95e	95	5	Dense	22806 (1995)	12232	5.35	134.2
B55e	55	5	Baseline	24421 (1395)	9232	5.80	247.9
B95e	95	5	Baseline	39676 (1995)	15112	7.95	811.3
S55e	55	5	Sparse	34311 (1395)	10442	7.08	5093.2
S95e	95	5	Sparse	56661 (1995)	17262	8.90	5524.0

Table 3.5: Model statistics for 12 problem instances.

We observe that the CPU time needed to solve an instance tends to grow rapidly when $|Q|$ increases, when $|S|$ increases, and when the network becomes sparser. This is well explained by the observations that, in each of these cases, the number of

variables and constraints increase and the LP relaxation of the MIP becomes weaker (i.e., the LP gap in the root node increases), which tends to negatively impact the performance of a branch-and-bound algorithm. Note that sparser networks bring about many additional variables in (3.22) and additional constraints in (3.17)-(3.20), because the number of RWC locations along a route, $|K_q|$, tends to increase when the network gets sparser.

3.5.4 Sensitivity Analysis

Proposition 3.2 proves worst-case bounds for the optimality gap induced by impreciseness in the demand parameters d_k and f_{qs}. To gain insight into the actual sensitivity, we randomly generate the "true" values of the parameters in the baseline case, and determine the resulting optimality gap for the balanced solution (see Section 3.5.1). Specifically, we draw for each of these parameters from the uniform distribution on the interval $[-\delta, \delta]$, represented by $U[-\delta, \delta]$, and determine the true values of f_{qs} and d_k as $f_{qs}^{\#} = f_{qs} \cdot (1 + U[-\delta, \delta])$ and $d_k^{\#} = d_k \cdot (1 + U[-\delta, \delta])$. We test different values of δ, and generate 50 demand scenarios for each of them. Table 3.6 shows for each value of δ the resulting average optimality gap, the maximum optimality gap, the non-parametric worst case bound on this gap (which is not tight for our specific instance) and the tight parametric bound. The latter is obtained by solving an MIP formulation of the problem defined in (3.16).

	δ				
Optimality gap	0.2	0.4	0.6	0.8	1.0
Avg. (%)	0.2	0.5	0.9	1.6	1.7
Max. (%)	0.5	1.2	2.7	2.9	3.8
Worst Case Par. (%)	2.5	4.2	10.7	26.6	36.9
Worst Case Nonpar. (%)	33.3	57.1	75.0	88.9	100.0

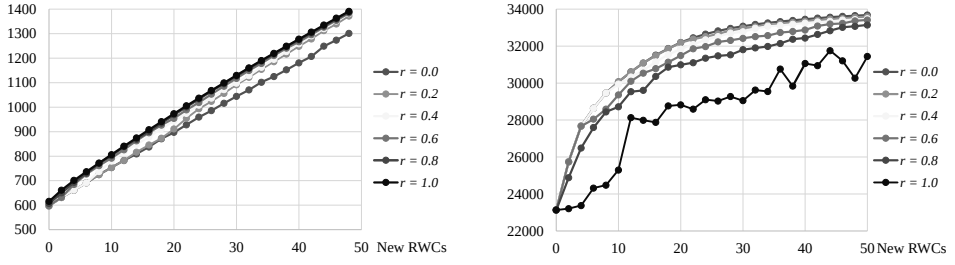
Note: Worst Case Par.: tight bound found by solving (3.16).
Worst Case Nonpar.: non-tight bound from Proposition 3.2.

Table 3.6: Optimality gap (% of opt.) by imprecise demand data.

The results indicate that the optimal solution is highly insensitive to impreciseness. Even when the true values of the demand parameters deviate by up to 100% from the presently used values, the maximum optimality gap attained in the 50 demand scenarios is only 3.8%. This can be explained by the fact that near-optimal solutions locate facilities along the busiest truck routes and at the busiest truck stops. The location decisions in these solutions therefore tend to overlap considerably. As a result, the effects of impreciseness are strongly correlated for these solutions, yielding only small differences between them. Nevertheless, worst-case analysis shows that a substantial optimality gap is possible.

3.5.5 Effect of the Number of New Facilities and Objective Weights.

Figures 3.7a and 3.7b summarize the results when solving the RHFLP for the baseline case with the number of facilities (with PC packages) to be placed $p \in \{2, 4, \ldots, 50\}$, the number of HC packages to be allocated $p_{HC} = \frac{p}{2}$, and relative importance parameter $r \in \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$. All other parameters are kept as before.



(a) Daily patient volume.

(b) Effectiveness score.

Figure 3.7: Daily patient volume and effectiveness score as a function of the number of new RWCs (p) and the relative importance of the patient volume criterion (r) and the the effectiveness criterion ($1 - r$).

The results show that focusing solely on one criterion (i.e., choosing $r = 0.0$ or $r = 1.0$) results in significant sub-optimality in terms of the other criterion. By setting $r = 0.8$ ($r = 0.2$), large gains can be made in terms of effectiveness (patient

volume) at a marginal loss in terms of patient volume (effectiveness). Notice that the $r = 0.4$ results are close to optimal for both patient volume and effectiveness, and that values lower than 0.4 apparently yield negligible further improvements in effectiveness. The "irregular behavior" of the effectiveness score for $r = 1.0$ is caused by the indifference on where to place the HC package, leading to more or less random package allocation decisions.

3.6 Discussion and Conclusions

Truck drivers play a key role in the spread of infectious diseases in Sub-Saharan Africa. Truck drivers thereby affect the health and lives of many, but also suffer from poor health and significantly reduced life expectancy themselves. Due to professional circumstances, their health service needs are generally not well addressed. This chapter considers the design problems arising when developing an integrated network of roadside healthcare facilities to meet the health service needs of the truck drivers, as operated by NGO North Star. More specifically we consider the problem of choosing optimal locations for clinics that are added to the existing network, and allocating (additional) service packages to clinics. The optimization considers two criteria, namely the number of truck drivers served and the effectiveness of the service delivery to the population served. The latter criterion is modeled through three novel access measures which capture the needs for effective health service provisioning.

The resulting optimization problem differs from previously studied healthcare facility location problems due to the specific mobile nature of health service demand of truck drivers. This chapter develops a mixed-integer linear programming model which captures the three access measures. We present computational analyses by solving problem instances arising from the network of major transport corridors in South-East Africa. Understanding the relationship between access to services and effectiveness is still in early stages of development, and availability of data is essentially limited in Sub-Saharan Africa. We therefore present a theoretical and computational analysis on the robustness of the model, investigating the relationship between data inaccuracy, solutions, and solution values. This analysis reveals that the solutions found are highly robust.

Until recently, the North Star NGO has predominantly made network design decisions based on the first objective criterion: maximize the number of truck drivers served. Our experiments indicate that considerable gains in effectiveness of the services are attainable when taking the second criterion into account, without having to reduce the number of truck drivers served much. We show that solutions that are close to optimal for both patient volume and effectiveness can be obtained by balancing the weights of the criteria in the objective function. Furthermore, our computational experiments reveal that the marginal increase in objective value when adding a facility or a service package to the network is not necessarily decreasing: synergy effects in terms of the effectiveness of service provisioning can be obtained when increasing the number of facilities that offer a given package along a truck route. Decision makers can exploit this property by making *long-term* investment plans for the network of RWCs instead of taking investment decisions sequentially.

Although, as we prove, the RHFLP is strongly $\mathcal{NP}$-hard, our numerical experiments show that the real life test instances can be solved within reasonable computation times using standard software. Further research on approximation and exact solution methods will be valuable if future applications consider more complex and larger problem instances. Another relevant direction is to (empirically) investigate the validity and the impact of the assumptions underlying our parameter estimates. For instance, our model implicitly assumes that patient volumes at the locations are unaffected by network design decisions. Likewise, models which include stochastic parameters can be of value to solve practical instances, given the many uncertainties in the data.

Our model disregards equity in health delivery, disadvantaging for instance truck drivers traveling along less frequently traveled routes. As access to health services is a human right laid down in the WHO constitution (United Nations, 1946), we advocate that, once the high volume needs are addressed, equity considerations should be investigated and taken into account in future work on network expansion. This gives rise to questions regarding the trade off between equity and effectiveness (see, e.g., Ares et al. (2016); McCoy and Lee (2014) for related work).

The proposed model captures the effectiveness of the services provided via the three newly proposed measures of access. There is much room for improvement

in the understanding of the relationships between access and effectiveness, and the role of the services provided (e.g., integrating HIV and TB care is known to bring about significant synergy effects). We believe that the models, and subsequently the solutions proposed and health services provided to the truck drivers can greatly benefit further understanding in this area.

Appendices

3.A Linearization of Effectiveness Variables e_{qs}

In order to model e_{qs} by means of a set of linear constraints, we now introduce additional notations and definitions related to the *the trip of a truck driver* in flow q (see Table 3.7 for an overview). For a given solution ω, we represent this trip by the following vector of vertices: $\pi_{qs} = [\pi_{qs}(1), \pi_{qs}(2), \ldots, \pi_{qs}(n-1), \pi_{qs}(n)]$. Here, $\pi_{qs}(1)$ and $\pi_{qs}(n)$ are the start vertex O_q and the end vertex D_q, respectively. The vector $[\pi_{qs}(2), \ldots, \pi_{qs}(n-1)]$ is the sequence of *package s locations* that are passed during a trip from O_q to D_q. Recall that for a given solution ω a package s location is a location k with a facility that offers package s, and note that $\pi_{qs}(1)$ and $\pi_{qs}(n)$ are not a package s location. As motivated in the main text, we make the following assumption:

Assumption 3.1. *A truck driver cyclically travels from O_q to D_q to O_q to D_q et cetera, as follows:*

$$\pi_{qs}(1), \pi_{qs}(2), \ldots, \pi_{qs}(n-1), \pi_{qs}(n), \pi_{qs}(n-1), \ldots, \pi_{qs}(2), \pi_{qs}(1), \pi_{qs}(2), \ldots]$$

t_{kl}	travel time between two locations k and l
L_q	set of vertices in the path of flow q, $L_q = \{O_q \cup K_q \cup D_q\}$
L_{kq}	set of locations that are passed *after* passing location k during a trip from O_q to D_q
π_{qs}	set of vertices representing a trip from O_q to D_q = [O_q, package s locations along path, D_q]
i_{klqs}	$\begin{cases} 1 & \text{if location } l \text{ is the immediate successor of location } k \text{ in vector } \pi_{qs} \\ 0 & \text{otherwise .} \end{cases}$
tc^{CTL}_{kl}	total time a truck driver in flow q is covered in segment $k \to l, (k,l) \in K_q \times K_q$
$tc^{CTL}_{kD_q}$	total time a truck driver in flow q is covered in segment $k \to D_q \to k, k \in K_q$
$tc^{CTL}_{O_q l}$	total time a truck driver in flow q is covered in segment $l \to O_q \to l, l \in K_q$
tc^{RCTL}_{kl}	effective time a truck driver in flow q is covered in segment $k \to l, (k,l) \in K_q \times K_q$
$tc^{RCTL}_{kD_q}$	effective time a truck driver in flow q is covered in segment $k \to D_q \to k, k \in K_q$
$tc^{RCTL}_{O_q l}$	effective time a truck driver in flow q is covered in segment $l \to O_q \to l, l \in K_q$

Note: see Section 3.3.2 for the definition of the effective time covered during period $[t_1, t_2]$

Table 3.7: Table of Notations.

Let us consider the case that $|\pi_{qs}| \geq 3$ (i.e., there is at least one package s location along the path). Because of Assumption 3.1, we know that a truck driver in flow q always returns to $\pi_{qs}(1)$ after reaching $\pi_{qs}(n)$, and vice versa. We refer to the cycle from $\pi_{qs}(1)$ via $\pi_{qs}(n)$ to $\pi_{qs}(1)$ as the *flow q cycle*. This cycle could be regarded as a set of paths between successively passed package s locations. This set is given by: $\{\pi_{qs}(h) \to \pi_{qs}(h+1) | 2 \leq h \leq n-2\} \cup \{\pi_{qs}(n-1) \to \pi_{qs}(n) \to \pi_{qs}(n-1)\} \cup \{\pi_{qs}(h+1) \to \pi_{qs}(h) | 2 \leq h \leq n-2\} \cup \{\pi_{qs}(2) \to \pi_{qs}(1) \to \pi_{qs}(2)\}$. We refer to these paths as *package s cycle segments* from now on. The lenght of a cycle segment represents the travel time of the truck driver for that segment, and is obtained via the parameters t_{kl}.

Example. For sake of conciseness, we omit the subscripts q and s in this example. Figure 3.8 illustrates the path of a truck driver travelling from origin $\pi(1)$ to destination $\pi(4)$. During his trip, he passes two package s locations: $\pi(2)$ and $\pi(3)$. Hence, the set of package s cycle segments is given by $\{\pi(2) \to \pi(3), \pi(3) \to \pi(4) \to \pi(3), \pi(3) \to \pi(2), \pi(2) \to \pi(1) \to \pi(2)\}$. Consequently, the set of travel time intervals between successively passed package s locations are given by: $\{t_{\pi(2),\pi(3)}, t_{\pi(3),\pi(4)} +$

$t_{\pi(4),\pi(3)}, t_{\pi(3),\pi(2)}, t_{\pi(2),\pi(1)} + t_{\pi(1),\pi(2)}\}$.

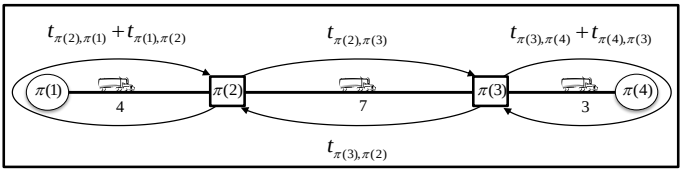


Figure 3.8: Package **s** cycle segments within the flow **q** cycle.

Let variable i_{klqs} be equal to 1 if location l is the immediate successor of location k in vector π_{qs} and equal to 0 otherwise. In Appendices 3.B, 3.C, and 3.D we prove the following results:

Proposition 3.4. *Constraints* (3.17) - (3.21) *ensure that variable* i_{klqs} *is equal to* 1 *if and only if location* l *is the immediate successor of location* k *in vector* π_{qs}.

$$\sum_{l \in L_q} i_{klqs} = y_{ks} \qquad q \in Q, k \in K_q, s \in S \tag{3.17}$$

$$\sum_{l \in L_q} i_{klqs} = 1 \qquad q \in Q, k \in O_q, s \in S \tag{3.18}$$

$$\sum_{k \in L_q} i_{klqs} = y_{ls} \qquad q \in Q, l \in K_q, s \in S \tag{3.19}$$

$$\sum_{k \in L_q} i_{klqs} = 1 \qquad q \in Q, l \in D_q, s \in S \tag{3.20}$$

$$i_{klqs} \in \{0,1\} \quad q \in Q, k \in L_q, l \in L_{kq}, s \in S \tag{3.21}$$

Lemma 3.1. *For solution* ω, *the polyhedron* $P_B(\omega) = \{\mathbf{i} | 0 \leq \mathbf{i} \leq 1, (3.17) - (3.20)\}$ *is a singleton.*

Lemma 3.2. *The constraint matrix* B *defined by the left-hand side of constraints* (3.17)-(3.20) *is totally unimodular (TU).*

Note that, for a given a solution ω, the right-hand side of constraints (3.17)-(3.20) is integral. It follows from the unimodularity of B that the polyhedron $P_B(\omega)$ is integral as well (Schrijver, 1998). By Lemma 3.1 we know that this polyhedron consists of *one* feasible solution, which must hence be integral. This shows that the solution space remains the same when defining variables i_{klqs} as *continuous* variables instead of binary variables:

$$i_{klqs} \in [0,1] \qquad q \in Q, k \in L_q, l \in L_{kq}, s \in S \tag{3.22}$$

Variables i_{klqs} provide a complete description of the package s cycle segments. Let us consider a given segment $k \to l$ for which $i_{klqs} = 1$. For this cycle segment the average access time is given by $\frac{1}{2}t_{kl}^2$, and the the total time covered tc_{kl}^{CTL} and the effective time covered tc_{kl}^{RCTL} (see equation (3.3)) are calculated as:

$$tc_{kl}^{CTL} = \min\left\{t_{kl}, \tau_s^{CTL}\right\}$$

$$tc_{kl}^{RCTL} = \begin{cases} t_{kl} & \text{if } t_{kl} < \tau_{1s}^{RCTL} \\ t_{kl} - \frac{1}{2}\frac{\left(t_{kl} - \tau_{1s}^{RCTL}\right)^2}{\tau_{2s}^{RCTL} - \tau_{1s}^{RCTL}} & \text{if } \tau_{1s}^{RCTL} \le t_{kl} \le \tau_{2s}^{RCTL} \\ \tau_{2s}^{RCTL} - \frac{1}{2}\left(\tau_{2s}^{RCTL} - \tau_{1s}^{RCTL}\right) & \text{if } t_{kl} > \tau_{2s}^{RCTL} \end{cases}$$

For segments $k \to D_q \to k$ and $l \to O_q \to l$ the values are obtained by replacing t_{kl} in these definitions by their durations: $t_{kD_q} + t_{D_qk}$ and $t_{O_ql} + t_{lO_q}$, respectively. Using these parameters and the variables i_{klqs}, the value of access measure $a_{qs}^j(\mathbf{y})$ is calculated by auxiliary (linear) constraints (3.23) - (3.25). For conciseness, we refer to this variable as a_{qs}^j from now on. The idea behind these constraints is that they calculate the value of a_{qs}^j per package s cycle segment and aggregate these in the correct way (the aggregation for a_{qs}^{CTL} and a_{qs}^{RCTL} is obvious, the aggregation for a_{qs}^{ASAP} is explained in De Vries (2011)):

$$a_{qs}^{CTL} = \frac{1}{T_q}\left(\sum_{l\in K_q} i_{O_q lqs} tc_{O_q l}^{CTL} + \sum_{k\in D_q} i_{kD_q qs} tc_{kD_q}^{CTL} + \sum_{k\in K_q}\sum_{l\in K_q} i_{klqs}\left(tc_{kl}^{CTL} + tc_{lk}^{CTL}\right)\right) \tag{3.23}$$

$$a_{qs}^{RCTL} = \frac{1}{T_q}\left(\sum_{l\in K_q} i_{O_q lqs} tc_{O_q l}^{RCTL} + \sum_{k\in D_q} i_{kD_q qs} tc_{kD_q}^{RCTL} + \sum_{k\in K_q}\sum_{l\in K_q} i_{klqs}\left(tc_{kl}^{RCTL} + tc_{lk}^{RCTL}\right)\right) \tag{3.24}$$

$$a_{qs}^{ASAP} = \frac{1}{2T_q}\left(\sum_{l\in K_q} i_{O_q lqs}\left(t_{O_q l} + t_{lO_q}\right)^2 + \sum_{k\in D_q} i_{kD_q qs}\left(t_{kD_q} + t_{D_q k}\right)^2 + \sum_{k\in K_q}\sum_{l\in K_q} i_{klqs}\left(t_{kl}^2 + t_{lk}^2\right)\right) + i_{O_q D_q qs} M \tag{3.25}$$

Here, M is a large number that sets a_{qs}^{ASAP} to a large constant if service package s is not offered along the path of flow q. Namely, constraints (3.18) and (3.20) ensure that $i_{O_q D_q qs} = 1$ in this case.

Example (cont'd). Let the travel times in the path introduced in the example be $t_{\pi(1),\pi(2)} = t_{\pi(2),\pi(1)} = 4$, $t_{\pi(2),\pi(3)} = t_{\pi(3),\pi(2)} = 7$, $t_{\pi(3),\pi(4)} = t_{\pi(4),\pi(3)} = 3$. This implies that $T = 28$. Let the critical time-limit τ be equal to 5. The light arcs in Figure 3.9 illustrate the time slots during which a truck driver is covered. For each of the segments, the duration of this time-slot equals 5. Therefore $a^{CTL} = \frac{20}{28}$. Next, let τ_1^{RCTL} and τ_2^{RCTL} be equal to 2 and 3, respectively. Figure 3.10 illustrates the degree of coverage for each part of the cycle: it equals 1 for parts in which the travel time to the next package s location is at most 2, and decreases to 0 when this travel

time increases to 3. The effective time covered equals 2.5 for each cycle segment. Consequently, a^{RCTL} equals $\frac{10}{28}$. Finally, the average value of $\Delta(t)$ during a trip along the cycle is given by $a^{ASAP} = \frac{1}{2 \cdot 28}\left((4+4)^2 + (3+3)^2 + 7^2 + 7^2\right) \approx 3.5$.

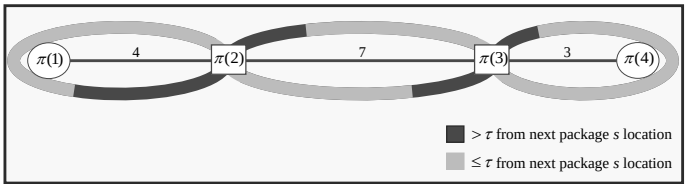


Figure 3.9: Time-slots during which truck drivers are at most 5 time-units from the next package **s** location.

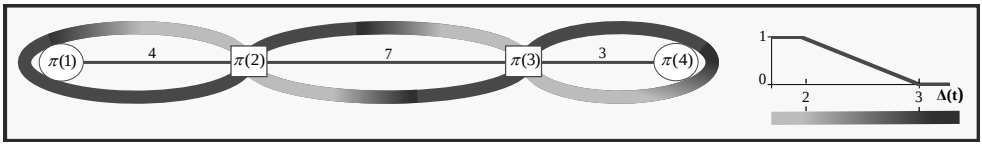


Figure 3.10: Left: degree of coverage for each part of the cycle, with recommended time-limit 2 and critical time-limit 3. Right: degree of coverage as a function of the access time.

Let M_{qs} represent an upper bound on a^j_{qs}. The set of auxiliary (linear) constraints that model e_{qs} as a piecewise linear function of a^j_{qs} can be formulated using the so-called Lambda Method (Lee and Wilson, 2001):

$$e_{qs} = \lambda_{3qs} + \lambda_{4qs} \quad q \in Q, s \in S_j \quad j \in \{CTL, RCTL\} \tag{3.26}$$

$$e_{qs} = \lambda_{1qs} + \lambda_{2qs} \quad q \in Q, s \in S_j \quad j = ASAP \tag{3.27}$$

$$\lambda_{1qs} + \lambda_{2qs} + \lambda_{3qs} + \lambda_{4qs} = 1 \quad q \in Q, j \in J, s \in S_j \tag{3.28}$$

$$\lambda_{1qs} 0 + \lambda_{2qs} \underline{\alpha}_s^j + \lambda_{3qs} \overline{\alpha}_s^j + \lambda_{4qs} M_{qs} = a_{qs}^j \quad q \in Q, j \in J, s \in S_j \tag{3.29}$$

$$\lambda_{1qs} \leq z_{1qs} \quad q \in Q, j \in J, s \in S_j \tag{3.30}$$

$$\lambda_{2qs} \leq z_{1qs} + z_{2qs} \quad q \in Q, j \in J, s \in S_j \tag{3.31}$$

$$\lambda_{3qs} \leq z_{2qs} + z_{3qs} \quad q \in Q, j \in J, s \in S_j \tag{3.32}$$

$$\lambda_{4qs} \leq z_{3qs} \quad q \in Q, j \in J, s \in S_j \tag{3.33}$$

$$z_{1qs} + z_{2qs} + z_{3qs} = 1 \quad q \in Q, j \in J, s \in S_j \tag{3.34}$$

$$\lambda_{iqs} \geq 0 \quad i \in \{1, 2, 3, 4\} \tag{3.35}$$

$$z_{iqs} \in \{0, 1\} \quad i \in \{1, 2, 3\} \tag{3.36}$$

Summarizing, to transform model (3.8) - (3.15) into a mixed-integer linear programming model, it suffices to replace equations (3.9) by constraints (3.17) - (3.20) and (3.22) - (3.36).

3.B Proof of Correctness of Constraints (3.17)-(3.21)

Proof. Let us consider a given flow q and package s. We refer to a location l for which $i_{klqs} = 1$ as a successor of k and to a location k for which $i_{klqs} = 1$ as a predecessor of l. We claim that the constraints induce the $n-1$ successor-predecessor pairs $(\pi_{qs}(h), \pi_{qs}(h+1)), h = 1, ..., n-1$, and that $i_{\pi_{qs}(h),\pi_{qs}(h+1)qs} = 1$ for $h = 1, ..., n-1$, as required.

We prove our claim by induction. Note that, by definition of the variables i_{klqs}, we know that a condition for location l to be a successor of k is that it is visited *after* location k during a trip along the path of flow q, as required. Note further

that, if a location k has exactly one successor l, then constraints (3.17) and (3.18) ensure that $i_{klqs} = 1$. By definition of the variables i_{klqs}, we know that $\pi_{qs}(n)$ has no successor. Constraint (3.20) ensures that $\pi_{qs}(n)$ has at least one predecessor, and constraint (3.17) ensures that $\pi_{qs}(n-1)$ has at least one successor. Hence, this successor must be $\pi_{qs}(n)$, forcing that $i_{\pi_{qs}(n-1),\pi_{qs}(n)qs} = 1$. Hence, our claim holds for $h = n-1$. Now, we show by induction that our claim holds for $h-1$, given that it holds for $h, ..., n-1$. Constraint (3.17) imposes that $\pi_{qs}(h-1)$ has at least one successor. This implies that there exists an index $h' > h-1$ such that $\pi_{qs}(h')$ has $h-1$ as a predecessor. By our induction hypothesis, we know that each location $\pi_{qs}(\hat{h})$ for which $\hat{h} > h$ already has the location $\pi_{qs}(\hat{h}-1)$ as a predecessor, and that $i_{\pi_{qs}(\hat{h}-1),\pi_{qs}(\hat{h})qs} = 1$. Hence, none of the locations $\pi_{qs}(\hat{h})$ for which $\hat{h} > h$ can be a successor of $\pi_{qs}(h)$. Hence, $\pi_{qs}(h)$ must be the only successor of $\pi_{qs}(h-1)$, and $i_{\pi_{qs}(h-1),\pi_{qs}(h)qs} = 1$. This shows that our claim is true for $h = 1, ..., n-1$.

Finally, observe that each of the constraints is met if $i_{\pi_{qs}(h),\pi_{qs}(h+1)qs} = 1$ for the $n-1$ successor-predecessor pairs $(\pi_{qs}(h), \pi_{qs}(h+1))$. This implies that for the given q and s, no other variable i_{klqs} can attain a positive value, as required. □

3.C Proof of Unique Solution for Polyhedron $P_B(\omega)$

Proof. This follows from the same argument as the one used in Appendix 3.B. Consider a given flow q and a given package s. We now refer to a location l for which $i_{klqs} > 0$ as a successor of k and to a location k for which $i_{klqs} > 0$ as a predecessor of l. Using induction it follows that the constraints allow for one unique solution: they induce the $n-1$ successor-predecessor pairs $(\pi_{qs}(h), \pi_{qs}(h+1)), h = 1, ..., n-1$, and $i_{\pi_{qs}(h),\pi_{qs}(h+1)qs} = 1$ for $h = 1, ..., n-1$. □

3.D Proof of Total Unimodularity of the Constraint Matrix Corresponding to Constraints (3.17)-(3.20)

Proof. Let us represent the constraint matrix corresponding to constraints (3.17)-(3.20) by B, which has one column for each variable i_{klqs}. Next, we prove that B meets conditions that are sufficient for this matrix to be totally unimodular, as specified in (Heller and Tompkins, 1956).

Note that every entry in B is a 0 or 1. Let us write matrix B as $[B_1; B_2]$, where B_1 represents the matrix formed by the rows corresponding to the "outflow constraints" (3.17) and (3.18), and where B_2 represents the matrix formed by the rows corresponding to the "inflow constraints" (3.19) and (3.20). Consider a given service package s, a given truck flow q and a given arc (k, l) for which the variable i_{klqs} is defined. Constraints (3.17) and (3.18) ensure that in the corresponding column in B_1 there is precisely one entry with the value 1. Similarly, constraints (3.19) and (3.20) ensure that in the corresponding column in B_2 there is precisely one entry with the value 1. This shows that we can partition the rows of B into two disjoint sets forming the matrices B_1 and B_2 such that, if two non-zero entries in a column of B have the same sign, then the one entry is in B_1 and the other row is in B_2, as required. As implied by the previous observation, every column contains at most two non-zero entries. □

3.E Proof of Strong $\mathcal{NP}$-hardness of RHFLP

Proof. The fact that the decision version of RHFLP is in $\mathcal{NP}$ follows from the observation that a YES answer can be verified in polynomial time through the MIP formulation. In what follows, we propose a polynomial time reduction from CLIQUE – a strongly $\mathcal{NP}$-Complete (Papadimitriou and Steiglitz, 1998) problem – to RHFLP's decision problem. This implies the $\mathcal{NP}$-Completeness of the latter.

CLIQUE represents the following decision problem: "does graph $G(V, E)$ contain a complete sub-graph consisting of k nodes?" The transformation introduces a po-

tential RWC location $k \in KP$ for each node $v \in V$ and a path $q \in Q$ for each edge $e \in E$. Path q introduced for edge (v_1, v_2) is a direct path between the potential RWC locations introduced for nodes v_1 and v_2. Furthermore, we set $|S| = 1$, $p = p_s = k$, $f_{qs} = 1 \forall q$, $d_k = 0$, $w_s = 1$, and $t_{kl} = 1 \quad \forall (k, l)$. Other parameters values are chosen so as to make the effectiveness of package s provisioning for flow q equal 1 if and only if RWCs are placed at both endpoints of the corresponding path. If $j = CTL$, this is realized when $\tau_s^{CTL} = 1$, $\underline{\alpha}_s^{CTL} = 1$, $\overline{\alpha}_s^{CTL} = 1$. If $j = RCTL$, it suffices to set $\tau_{1s}^{RCTL} = 1$, $\tau_{2s}^{RCTL} = 1$, $\underline{\alpha}_s^{RCTL} = 1$, $\overline{\alpha}_s^{RCTL} = 1$. Finally, if $j = ASAP$, values $\underline{\alpha}_s^{ASAP} = 0.5$, $\overline{\alpha}_s^{ASAP} = 0.5$ suffice. The fact that $|KP| = |V|$ and $|Q| = |E|$ makes this a polynomial time reduction.

We shall now show that CLIQUE has a YES-answer if and only if the solution value of the corresponding RHFLP decision problem instance equals $k \cdot (k-1)/2$.

If. If the RHFLP instance has a value of $k \cdot (k-1)/2$, the number of flows for which the effectiveness of package s provisioning is strictly positive equals *at least* $k \cdot (k-1)/2$. The fact that this number is *at most* $k \cdot (k-1)/2$ follows from the fact that service packages s can only be allocated to k RWC locations. Hence, this number must be *exactly* $k \cdot (k-1)/2$, which is only the case if exactly $k \cdot (k-1)/2$ unique flow paths have the k allocated service packages as end points. Hence, the corresponding CLIQUE instance must have a clique of size k.

Only if. At most $k \cdot (k-1)/2$ flow paths have a strictly positive effectiveness of package s provisioning since we allocate service packages s to k RWC locations. Further, the solution value *equals* the number of flow paths for which effectiveness of the package s provisioning per truck driver equals 1, since effectiveness equals either 1 or 0. Combining these observations implies the solution value equals *at most* $k \cdot (k-1)/2$. If $G(V, E)$ has a$k-$clique, the corresponding k RWC locations connect exactly $k \cdot (k-1)/2$ flow paths. Effectiveness of package s provisioning for each of these flows equals 1 when allocating package s to these k locations. Hence, the solution value of the RHFLP is *exactly* $k \cdot (k-1)/2$. □

3.F Proof of Proposition 3.2

Proof. By definition, $v^*(\omega^*) \geq v^*(\omega^\#)$. In addition, the objective function is linear in f_{qs} and d_k, which implies that $v^\#(\omega^*) \geq (1-\delta)v^*(\omega^*)$ and that $v^\#(\omega^\#) \leq (1+\delta)v^*(\omega^\#)$. Hence, $\frac{v^\#(\omega^*)}{1-\delta} \geq v^*(\omega^*) \geq v^*(\omega^\#) \geq \frac{v^\#(\omega^\#)}{1+\delta}$.

This worst case bound is attained under three conditions. First, $v^*(\omega^*) = v^*(\omega^\#)$. Second, $K^* \cap K^\# = \emptyset$ and $Q^* \cap Q^\# = \emptyset$, where $K^* = \{k|x_k = 1, \omega^*\}$, $K^\# = \{k|x_k = 1, \omega^\#\}$, $Q^* = \{q|e_{qs} > 0, s \in S, \omega^*\}$ and $Q^\# = \{q|e_{qs} > 0, s \in S\omega^\#\}$. Generally this second condition is not met, since it is likely that the solutions ω^* and $\omega^\#$ both locate facilities at the busiest truck stops (i.e., $K^* \cap K^\# \neq \emptyset$), and provide the busiest truck flows with continuous access to service package s (i.e., $Q^* \cap Q^\# \neq \emptyset$). Third, let $f^\#_{qs}$ and $d^\#_k$ denote the true values of the parameters f_{qs} and d_k. Then $f^\#_{qs} = (1+\delta)f_q \forall q \in Q^\#$, $f^\#_{qs} = (1-\delta)f_q \forall q \in Q^*$, $d^\#_k = (1+\delta)d_k \forall k \in K^\#$, $d^\#_k = (1-\delta)d_k \forall k \in K^*$. Under these conditions, $v^\#(\omega^*) = (1-\delta)v^*(\omega^*)$ and $v^\#(\omega^\#) = (1+\delta)v^*(\omega^\#)$. Rewriting shows that the bound is attained with equality. □

3.G Proof of Proposition 3.3

Proof. We rewrite the objective function as:

$$rZ_{PV} + (1-r)\sum_{s\in S} w_s \hat{Z}_s = \sum_{ob\in OB} w_{ob} Z_{ob} \tag{3.37}$$

Here, for package s and service package type j, $\hat{Z}_s = \sum_{q\in Q} f_{qs}\hat{g}^j_s(a^j_{qs})$, where $\hat{g}^j_s(\cdot) = \frac{g^j_s(\cdot)}{w_s}$. Furthermore, let $w^\#_{ob}$ and w^*_{ob} denote "optimal" and the presently used value for w_{ob}, respectively, and let $Z^\#_{ob}$ and Z^*_{ob} denote the value Z_{ob} for solutions $\omega^\#$ and ω^*, respectively. We rewrite $v^\#(\omega^\#) - v^\#(\omega^*)$ to $\sum_{ob\in OB^+} w^\#_{ob}\left(Z^\#_{ob} - Z^*_{ob}\right) + \sum_{ob\in OB^-} w^\#_{ob}\left(Z^\#_{ob} - Z^*_{ob}\right)$. Here, $OB^+ = \{ob|Z^\#_{ob} > Z^*_{ob}\}$ and $OB^- = \{ob|Z^\#_{ob} \leq Z^*_{ob}\}$. Note that $v^\#(\omega^\#) - v^\#(\omega^*)$ is maximized if $w^\#_{ob} = (1-\delta)w^*_{ob}$ for all $ob \in OB^-$ and if $w^\#_{ob} = (1+\delta)w^*_{ob}$ for all $ob \in OB^+$. Next, observe that the fact that $v^*(\omega^*) \geq v^*(\omega^\#)$ implies $\sum_{ob\in OB^+} w^*_{ob}\left(Z^\#_{ob} - Z^*_{ob}\right) \leq -\sum_{ob\in OB^-} w^*_{ob}\left(Z^\#_{ob} - Z^*_{ob}\right)$.

Consequently,

$$v^{\#}(\omega^{\#}) - v^{\#}(\omega^{*}) \leq (1+\delta) \sum_{ob \in OB^{+}} w^{*}_{ob} \left(Z^{\#}_{ob} - Z^{*}_{ob} \right) \tag{3.38}$$

$$+(1-\delta) \sum_{ob \in OB^{-}} w^{*}_{ob} \left(Z^{\#}_{ob} - Z^{*}_{ob} \right) \tag{3.39}$$

$$\leq 2\delta \sum_{ob \in OB^{-}} w^{*}_{ob} \left(Z^{*}_{ob} - Z^{\#}_{ob} \right) \tag{3.40}$$

$$\leq 2\delta \sum_{ob \in OB} w^{*}_{ob} \left(Z^{*}_{ob} - \underline{Z}_{ob} \right) \tag{3.41}$$

Here, $\underline{Z}_{ob}$ denotes a lower bound on Z_{ob} for any solution ω (e.g., the value of Z_{ob} in the network without the new facilities).

Using the fact that $v^{*}(\omega^{*}) \leq \frac{v^{\#}(\omega^{*})}{1-\delta}$, the bound can be rewritten to:

$$v^{\#}(\omega^{\#}) \leq \frac{1+\delta}{1-\delta} v^{\#}(\omega^{*}) - 2\delta \sum_{ob \in OB} w^{*}_{ob} \underline{Z}_{ob} \tag{3.42}$$

$$= \frac{1+\delta}{1-\delta} v^{\#}(\omega^{*}) - 2\delta\gamma v^{*}(\omega^{*}) \tag{3.43}$$

$$\leq \frac{1+\delta}{1-\delta} v^{\#}(\omega^{*}) - 2\delta \frac{\gamma}{1+\delta} v^{\#}(\omega^{\#}) \tag{3.44}$$

Parameter γ is defined as (note: $\gamma \leq 1$):

$$\gamma = \frac{\sum_{ob \in OB} w^{*}_{ob} \underline{Z}_{ob}}{v^{*}(\omega^{*})} \tag{3.45}$$

By rewriting this inequality, we obtain the bound given in Proposition 3.3. The following example shows the tightness of the bound. Consider an instance for which $\gamma = 0$ and solutions $\omega^{\#}$ and ω^{*} such that $v^{*}(\omega^{*}) = v^{*}(\omega^{\#})$, $OB^{*} \cup OB^{\#} = \emptyset$, where $OB^{*} = \{ob | Z^{*}_{ob} > 0\}$, and $OB^{\#} = \{ob | Z^{\#}_{ob} > 0\}$. Next, let $w^{\#}_{ob} = (1+\delta) w_{ob} \forall ob \in OB^{\#}$, and let $w^{\#}_{ob} = (1-\delta) w_{ob} \forall ob \in OB^{*}$. Then $v^{\#}(\omega^{*}) = (1-\delta) v^{*}(\omega^{*})$ and $v^{\#}(\omega^{\#}) = (1+\delta) v^{*}(\omega^{\#})$. Rewriting shows that the bound is attained with equality. □

Chapter 4

A Column Generation Approach for Locating Roadside Clinics in Africa based on Effectiveness and Equity[1]

4.1 Introduction

Long distance truck drivers in Sub-Saharan Africa are extremely vulnerable to HIV and other infectious diseases (Apostolopoulos and Sönmez, 2007). The underlying determinants seem to be loneliness, long separation from home, monotony, and stress, which make them engage in high-risk sexual behaviors (Morris and Ferguson, 2007).

[1]Apart from several minor adaptations, this chapter is a direct copy of the article: J. Nunez Ares, H. de Vries, D. Huisman. *A Column Generation Approach for Locating Roadside Clinics in Africa based on Effectiveness and Equity.* Published in European Journal of Operations Research.

This context thereby brings about huge health (and thereby social and economic) risks to this population, and seems to fuel the spread of HIV and other infectious diseases (Gatignon and Van Wassenhove, 2008).

Providing long distance truck drivers with continuous access to basic (HIV) prevention, treatment, and care services is therefore believed to be an effective way to combat HIV and its consequences. Traditional healthcare facilities, however, are generally incapable of delivering these services. The main reasons are that truck drivers do not have time or permission to deviate from their routes, that these facilities cannot be accessed by truck, and that the opening hours are inconvenient to the drivers (Gatignon and Van Wassenhove, 2008; Ferguson and Morris, 2007).

Non-governmental organization North Star Alliance (North Star) aims to fill this gap by placing so-called Roadside Wellness Centers (RWCs) at busy truck stops along major truck routes in Sub-Saharan Africa. These RWCs provide basic healthcare services, such as clinical services, HIV testing and counseling, and behavior change communication to truck drivers and surrounding populations. North Star's current network consists of 35 RWCs, and will expand considerably in the next couple of years. This is expected to bring about large health benefits to truck drivers and surrounding populations. In the first place, by placing new RWCs at busy truck stops, North Star makes sure that many truck drivers are provided with (at least a basic level of) access to the most needed health services when they spend the night there. In addition, establishing a dense network of RWCs ensures that many truck drivers also have continuous access to the needed health services. That is, they are sufficiently close to an RWC at every moment during their trip, which De Vries et al. (2014a) argue to be a requirement for several health services to be effective.

Choosing the locations of a given number of new RWCs presents novel and complex optimization problems. Decision makers face a huge number of possible location decisions, and have to balance multiple objectives. De Vries et al. (2014b) propose a mixed-integer programming (MIP) formulation for the problem to locate a given number of new RWCs and to decide on the set of health services to be offered by these RWCs. The objectives are to maximize the impact of the new RWCs in terms of the patient volume served and in terms of health service effectiveness, as determined by the extent to which the truck drivers have continuous access to these services.

Numerical and analytical results suggest that the model potentially yields significant improvement for the location decisions taken by North Star.

In spite of that, further improvement of the model is needed in two directions. First, as many other non-profit organizations, North Star operates in a complex environment that is characterized by many different objectives and stakeholders. As North Star's network grows larger, a third objective is gaining importance among them: to provide equitable access to healthcare. More specifically, inequalities in access among truck drivers using different truck routes are to be minimized, both for ethical reasons, as inspired by the right to health (United Nations, 1946), as well as for medical and financial reasons.

The second direction for improvement deals with the complexity of the location problem. De Vries et al. (2014b) show that the location problem is strongly $\mathcal{NP}$-complete, and numerical experiments show that solving large problem instances becomes extremely difficult. Moreover, the complexity of the problem will be increased considerably by including an equity criterion. Alternative model formulations and solution methods are required to deal with this.

This chapter considers these two directions in the context of the problem of *locating a given number of RWCs* (i.e., we do not consider the decisions on the health services to offer at the RWCs). Our contributions are fourfold. First, we introduce and motivate the equity criterion, and propose several measures for (in)equality in access to healthcare among mobile patient populations. Second, we propose and analyze a novel set partitioning type of formulation for this type of facility location problem. The strengths of this formulation are that it allows for a variety of objective functions (e.g., maximizing patient volume, ensuring continuous access, and providing equity), and that the integrality gap is very small. Third, to deal with the exponential number of variables our formulation brings about, we propose and analyze a column generation approach to solve it. Moreover, we investigate several strategies to speed up our algorithm, including dual stabilization, column pool management, and accelerated pricing. The latter solves the pricing problem to near-optimality as a sequence of shortest path problems. Last, we numerically assess the trade-off between the equity criterion and North Star's current criteria (patient volume and continuous access) based on randomly generated instances. Our results suggest that

solutions that are close to optimal with respect to each of the optimization criteria are attainable.

Though we specifically focus on the problem to select locations for a given number of RWCs, these contributions also apply to the extension in which decisions on the health services to offer are considered. Additionally, our contributions are also applicable to a variety of related facility location problems that deal with moving demand units. Examples include the positioning of refueling stations, billboards, detection or inspection stations, convenience stores, and ambulances.

The remainder of this chapter is organized as follows. Section 4.2 gives an overview of the relevant literature. Section 4.3 describes the problem and the optimization criteria in detail. Next, we present the set partitioning type of formulation in Section 4.4. Section 4.5 describes our basic column generation approach. After this, we describe our acceleration techniques in Section 4.6. Section 4.7 presents our results. This is followed by Section 4.8, which investigates the trade-off between equity and North Star's current objectives using a real life case study about the North South Corridor Network. Finally, in Section 4.9 we draw some conclusions and discuss possible directions for future research.

4.2 Literature Review

4.2.1 Equity

Equity involves the comparison of two or more populations (or individuals) along some dimension. To account for the equity criteria in facility location decisions, one needs to clarify the exact meaning of equity. Though literature on equity stresses the importance of the subject, particularly in a resource allocation context, there is no consensus on its definition (Waters, 2000). Young (1995) classifies resource allocation rules based on three equity concepts: parity (claimants should be treated equally), proportionality (goods should be divided in proportion to differences among claimants), and priority (the person with the greatest claim to the good should get it). Depending on the weights assigned to these concepts, many views on equity are possible, ranging from a totally egalitarian perspective (inequalities are unaccept-

able) to a Marxist perspective (inequalities should represent differences in need) to a Rawlsian perspective (inequality is only allowed if it benefits those least advantaged) (Williams and Cookson, 2000).

Equity measures compare *effects* of actions on different groups, and possibly weigh such effects based on the characteristics of a group (e.g., needs or size). An abundance of literature proposing and analyzing such measures is available. Marsh and Schilling (1994) provide a list of 20 equity measures that have been developed in the facility location context. Though each of these measures prefer a completely equitable distribution over any other distribution, they differ in their valuations of inequitable distributions. More specifically, they assign different weights to the concepts of parity, proportionality, and priority. Ogryczak (2000) studies the trade-off between effectiveness and equity corresponding to several equity metrics for location problems. In a later paper, Ogryczak (2009) investigates how to include equity measures while avoiding inferior solutions in terms of effectiveness (e.g. distances).

Literature reveals a wide variety of concepts and variables that are used in studies on equity. Concepts that are being used to compare the service or utility a health system provides to different populations include health status, distribution of resources, expenditures, utilization, and access (Goddard and Smith, 2001; Culyer and Wagstaff, 1993; Musgrove, 1986). For example, Cardoso et al. (2016) incorporate *equity of access, geographical equity, and socioeconomic equity* in long-term care and network design decisions. Griffin (2012) analyzes allocation strategies that aim at achieving equity in *health improvement* among different communities. McCoy and Lee (2014) consider equitable allocation strategies for *motor cycle trips* facilitating access to healthcare in rural areas. Besides the variety in concepts and variables, there also exist differences in the methods to include them in a model. For example, Cardoso et al. (2016) incorporate equity by enforcing targets to be met whereas Cardoso et al. (2015) incorporate equity measures in the objective function.

Given the amount of literature on equity, we can conclude that equity is a frequently applied and highly relevant optimization criterion. Ongoing discussions on the definition and measurement of equity have yielded in a wide variety of ways to incorporate this concept in optimization models, some of which we will discuss in the next section.

4.2.2 Facility Location and Access to Healthcare

Our work focuses on allocating resources to different populations of truck drivers and improving levels of access to health care. These are key objectives in the context of global health problems, which typically deal with scarcities of resources, and have therefore received much more attention from the Operations Research community. For example, Brandeau et al. (2003) analyze the optimal allocation of epidemic control resources over multiple non-interacting populations, Deo et al. (2015) optimize the allocation of HIV diagnosing services to different health facilities, and McCoy and Johnson (2014) optimize the allocation of clinic capacity over time. White et al. (2011) have written an overview of Operations Research applications in low income countries. They show that research on relevant diseases in our context, in particular on HIV, is gaining more and more attention among the Operations Research community.

Since *access* to healthcare is continuously changing for mobile populations like truck drivers, De Vries et al. (2014b) introduced the concept of *continuous access* to healthcare. Continuous access provides truck drivers with adequate access at any point in time - which facilitates timely drug refills and prompt access in case of complications - and enables them to obtain services that require frequent clinic visits such as routine screening and treatment for HIV and TB. De Vries et al. (2014a) argue that this brings significant health benefits for truck drivers. One specific mechanism that explains the relationship between healthcare accessibility and health outcomes is treatment adherence. Especially for infectious diseases like TB and HIV, adherence is critical for treatment effectiveness and for reducing transmission and it is highly determined by accessibility (Munro et al. (2007), Mills et al. (2006)). In general, being able to receive continuous and coordinated care over time has been shown to lessen treatment delay and to stimulate treatment adherence, and thereby to decrease disease progression, disease transmission, and drug resistance (see e.g. Starfield et al., 2005).

Papers on healthcare facility location problems have been reviewed by Rahman and Smith (2000) and Daskin and Dean (2004). They show that these problems are typically solved by minimizing a function of the distances between facilities and

patients. For example, Smith et al. (2009) model the problem as a p-median problem, minimizing the (weighted) average distance to the nearest facility. Tanser (2006) and Verter and Lapierre (2002) go one step further by basing location decisions on an explicit model for the relationship between distance and healthcare utilization. The healthcare facility location models proposed or reviewed in these papers have one element in comon: they assume patients to be static. Since truck drivers are continuously changing position, these models are inapplicable in our context. The branch of *mobile* facility location problems does bear a close resemblance to our context. Similar to the problem to locate roadside clinics, these problems implicitly deal with determining the time intervals at which demand units can obtain certain health services (see e.g. Doerner et al., 2007).

Our problem can be classified as a network location problem. More specifically, it belongs to the class of flow interception problems, since locations of new facilities are optimized with respect to a flow of demand units. A flow of demand units in our work corresponds to a collection of truck drivers traveling the same route. Since such flow is not simply provided with sufficient access when only one RWC is placed along the corresponding route, we can further classify our problem as a multi-coverage flow interception problem.

Next to the highly related Roadside Healthcare Facility Location Problem introduced by De Vries et al. (2014b)(see Section 4.1), there are, as far as we know, only two other problems in the multi-coverage problems field, namely the billboard location problem (Averbakh and Berman, 1996) and the flow refueling location problem (Kuby and Lim, 2005). Our problem differs from the former because our degree of coverage is determined by the *driving* times between facilities rather than the *number* of facilities. Our problem differs from the latter in that the flows are covered to *some* extent in our problem, while in the refueling location problem they are either covered or not covered.

4.3 Problem Description

Before we describe our problem in detail, let us introduce some notations. Let KP denote the set of potential RWC locations – potential locations for new RWCs – and

let KC denote the set of current RWC locations – locations corresponding to the current RWCs. The union of these sets $KC \cup KP$ is the set of RWC locations K. Furthermore, let binary variable x_k indicate whether or not an RWC is available at a given RWC location $k \in KP$. We consider the problem to select locations for p new RWCs from the set KP, and measure the quality of a solution with respect to the patient volume criterion, the continuous access criterion, and the equity criterion by variables Z_{PV}, Z_{CA}, and Z_{EQ}, respectively. Finally, w_{PV}, w_{CA}, and w_{EQ} represent the relative importance of the three criteria. Then our objective function is defined as:

$$\max \quad w_{PV} Z_{PV} + w_{CA} Z_{CA} + w_{EQ} Z_{EQ} \tag{4.1}$$

The definitions of Z_{PV}, Z_{CA}, and Z_{EQ} use the following notations. Parameter d_k denotes the expected daily number of patients entering an RWC at location k. The set of long distance truck routes using the road network is denoted by Q and indexed by q. We measure the level of access along truck route q by means of the variable $c_q \in [0, 1]$, which we refer to as the *coverage score* of route q. Additionally, we relate to route q an ordered set of RWC locations $K_q \subseteq K$, where elements $K_q(1), K_q(2), ...$ represent the $1^{st}, 2^{nd}, ...$ RWC location passed when traveling the route. Similarly, KP_q and KC_q denote the ordered sets of potential and current RWC locations, respectively. Given location decisions $\mathbf{x} = \{x_k\}$, $K_q^{\mathbf{x}}$ represents the ordered set of RWC locations along route q for which $x_k = 1$. Finally, the number of truck drivers traveling route q is represented by f_q.

De Vries et al. (2014b) propose definitions of Z_{PV} and Z_{CA} based on interviews with North Star's CEO and staff. The next sections briefly describe these definitions and the rationale behind them, and introduce and motivate the equity criterion and corresponding measures.

4.3.1 Patient Volume

The patient volume criterion refers to the objective of serving as many truck drivers as possible. The rationale behind this objective is that the more truck drivers can be

provided with at least a *basic level of access* to the needed health services, the better it is. Namely, such basic level of access suffices to provide these truck drivers with many important "single shot" services such as HIV testing and counseling, behavioral change communication, and condom distribution.

Given location decisions $\mathbf{x}$, we measure patient volume Z_{PV} as the expected total daily number of patients entering an RWC:

$$Z_{PV} = \sum_{k \in KP} d_k x_k \tag{4.2}$$

The patient volume objective provides an incentive to place RWCs at *busy truck stops*, as many truck drivers can be reached there.

4.3.2 Continuous Access

Whereas providing truck drivers with a *basic level of access* suffices to effectively provide some health services, providing *continuous access* is beneficial or even required for effectively providing other health services, as argued in Section 4.2.2. Variable c_q, the coverage score of truck drivers traveling route q, indicates to what extent the level of access to healthcare provided along this route is "sufficient" or "effective", and ranges from 0 (insufficiently covered) to 1 (sufficiently covered). Given location decisions $\mathbf{x}$, we measure the *total* level of continuous access provided as the sum of the coverage scores of all truck drivers in the network, and refer to this measure as the continuous access score:

$$Z_{CA} = \sum_{q \in Q} f_q c_q \tag{4.3}$$

Before we formally define c_q, we make the following assumption:

Assumption 4.1. *RWC equivalents are accessible at the origin and destination of each truck route.*

An RWC equivalent refers to other healthcare facilities that also contribute to continuity of access along a truck route. As many truck route origins and destinations correspond to large cities, this assumption holds in most cases. We regard these facilities as RWCs in our model, implying that $K_q^{\mathbf{x}}(1)$ and $K_q^{\mathbf{x}}(m)$ correspond to current RWC locations at the origin and destination of route q, respectively (here, $m = |K_q^{\mathbf{x}}|$). Based on this assumption, we can split up the truck route in $m-1$ sub-routes between two adjacent RWCs along the route: $\left(K_q^{\mathbf{x}}(j), K_q^{\mathbf{x}}(j+1)\right), j \in \{1, 2, ..., m-1\}$, corresponding to a travel time of $t_{\left(K_q^{\mathbf{x}}(j), K_q^{\mathbf{x}}(j+1)\right)}$ time units.

Now consider a given moment during the trip of a truck driver along route q. We define him to be "safe" if the travel time to the next RWC along his route (i.e., his access time) is at most a given critical time-limit τ, and measure c_q as a piece-wise linear function $g(\cdot)$ of the fraction of time he is "safe" during his trip:

$$c_q = g\left(\frac{\sum_{j=1}^{m-1} \min\{t_{\left(K_q^{\mathbf{x}}(j), K_q^{\mathbf{x}}(j+1)\right)}, \tau\}}{T_q}\right) \tag{4.4}$$

Here, T_q denotes the total travel time for route q and the numerator represents the total time a truck driver traveling route q is safe. Figure 4.1 shows an example of the function $g(\cdot)$ with three different sections. For later use, we denote the two breakpoints of the function by μ_1 and μ_2, respectively, and represent the three segments of by $\sigma_1 = [0, \mu_1], \sigma_2 = [\mu_1, \mu_2]$, and $\sigma_3 = [\mu_2, 1]$. In this example, if the truck driver is safe less than 40% of the time, he is regarded as being insufficiently covered and his coverage score c_q equals 0. If he is safe 80% or more of the time, then we regard him as being sufficiently covered and his coverage score equals 1. Finally, his coverage score equals $\frac{e_q - 0.4}{0.8 - 0.4}$ if his fraction of time safe e_q is less than 80% but more than 40%. For example, his coverage score equals 0.5 if he is safe 60% of the time.

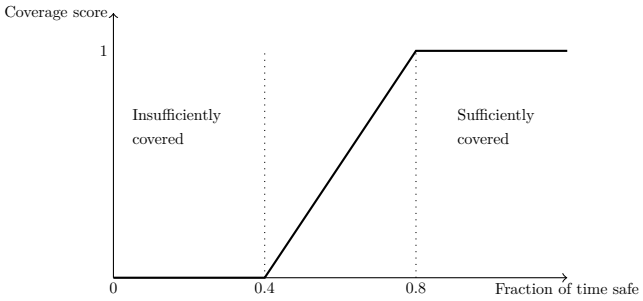


Figure 4.1: Coverage score as a piece-wise linear function of the fraction of time safe, with $\mu_1 = 0.4, \mu_2 = 0.8$.

4.3.3 Equity

The equity issue arises in a natural way in our healthcare facility location problem. Truck drivers generally drive along the same truck route for years (personal communication, North Star, 2011), so that, to a large extent, each truck route corresponds to a unique population of truck drivers. Different location allocation decisions have a different impact in terms of access to healthcare along these truck routes, so that these decisions implicitly determine the (in)equality in access to healthcare among the different populations.

As North Star's network is growing larger, this issue is gaining more and more importance. Whereas it can easily be defended why North Star should assign its first RWCs to the largest populations, one faces a dilemma as soon as the number of RWCs in the network grows. Namely, one has to decide whether to assign new RWCs to large and relatively well-served truck populations, where the total health impact may be largest, or to underserved populations, where the total health impact may be smaller. Hence, the principle of providing equal access to healthcare for truck drivers who (supposedly) have equal needs starts playing a role. Next to this ethical reason, there also is a pragmatic reason for taking the equity criterion into account: donors may be more willing to finance North Star's operations in case that a large part of the truck driver population benefits from them.

As mentioned in Section 4.2.1, equity measures compare *effects* of actions on different groups. In our context, the different groups are the truck driver populations corresponding to the different routes, the actions refer to location decisions, and the effects refer to coverage scores. Some of the most common equity measures (see Ogryczak, 2000; Marsh and Schilling, 1994) are the *sum of absolute deviations* (SAD), the *mean absolute deviation* (MAD), the *minimum effect* (ME), and the *Gini coefficient* (GC):

$$Z_{EQ}^{SAD} = \sum_{q_1,q_2 \in Q | q_2 > q_1} |c_{q_1} - c_{q_2}| f_{q_1} f_{q_2} \tag{4.5}$$

$$Z_{EQ}^{MAD} = \sum_{q \in Q} |c_q - \bar{c}| f_q \tag{4.6}$$

$$Z_{EQ}^{ME} = \min_{q \in Q} c_q \tag{4.7}$$

$$Z_{EQ}^{GC} = \frac{\sum_{q_1,q_2 \in Q | q_2 > q_1} |c_{q_1} - c_{q_2}| f_{q_1} f_{q_2}}{2 \sum_{q \in Q} c_q f_q} \tag{4.8}$$

Here, $\bar{c} = \frac{\sum_{q \in Q} c_q f_q}{\sum_{q \in Q} f_q}$. Measure Z_{EQ}^{ME} only considers the population that is least well-off, and thereby neglects inequalities among the other populations. The Gini coefficient Z_{EQ}^{GC}, which is a normalized variant of measure Z_{EQ}^{SAD}, in contrast, does consider the inequalities among all populations. However, this measure has the disadvantage to be highly non-linear, which makes the resulting optimization problem extremely complex. MeasureZ_{EQ}^{MAD} minimizes the sum of the absolute deviations from the mean, but neglects how these absolute deviations are distributed over the routes. This can lead to inferior solutions, as we show in the example presented below. Measure Z_{EQ}^{SAD} does indicate the spread of the outcomes (Ogryczak, 2009), and thereby avoids this. Hence, among the four proposed measures, we deem Z_{EQ}^{SAD}, a frequently used measure in facility location problems (Ogryczak, 2009, 2000; Marsh and Schilling, 1994), to be most suitable in our context.

Example 4.1. Figure 4.1 and Table 4.2 describe a small network with 3 truck routes and 3 potential RWC locations (locations 2, 3, and 6). We evaluate the effects

of adding one RWC to the network in terms of equity, as quantified by the four proposed equity measures. To calculate the coverage scores resulting from a location choice, we make use of the piece-wise linear function presented in Figure 4.1 (i.e., $\mu_1 = 0.4$ and $\mu_2 = 0.8$). Furthermore, we assume the critical time limit to be 60 hours and the truck volumes to be equal to 1. Intuitively, it is clear that the most equitable location choice would be to assign the RWC to location 2, as then both route 1 and route 2, which have equal access to healthcare in the initial situation, benefit from it. We will however show that measures Z_{EQ}^{ME} and Z_{EQ}^{MAD} regard all location choices to be equally attractive.

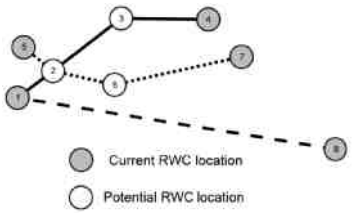


Figure 4.2: Example network.

Route	RWC locations	Travel times
1	1, 2, 3, 4	10, 50, 40
2	5, 2, 6, 7	10, 20, 70
3	1 ,8	200

Table 4.1: Route characteristics.

Table 4.2 describes the effects of the different location choices on the levels of access, as measured by the fraction of time safe, the corresponding coverage scores, and the resulting equity scores. Let us first illustrate the calculation of the coverage scores. When locating the RWC at location 3, then for route 1 there are two sub-routes between two adjacent RWCs: sub-route $(1, 3)$ and sub-route $(3, 4)$ with travel times 60 hours and 40 hours, respectively. In this situation, truck drivers traveling this route can reach a clinic within the critical time limit of 60 hours at each point of time during their trip. Hence, their fraction of time safe, calculated as $(\min\{60, 60\} + \min\{40, 60\})/100$, equals 1. Using the piece-wise linear function depicted in Figure 4.1, this translates into a coverage score of 1. For route 2, there is only one sub-route between adjacent RWCs: $(5, 7)$ with travel time 100 hours. Hence, the fraction of time safe, calculated as $\min\{100, 60\}/100$, equals 0.6, which corresponds to a coverage score of $(0.6 - 0.4)/(0.8 - 0.4) = 0.5$.

RWC at	Fract. safe	Coverage score	Z_{EQ}^{SAD}	Z_{EQ}^{MAD}	Z_{EQ}^{ME}	Z_{EQ}^{GC}
2	0.7/0.7/0.3	0.75/0.75/0	1.5	1	0	0.5
3	1/0.6/0.3	1/0.5/0	2	1	0	0.67
6	0.6/0.9/0.3	0.5/1/0	2	1	0	0.67

Table 4.2: Quality of location decisions in terms of equity, as measured by the four proposed equity measures.

Since none of the location choices improves the coverage score for the route that is least well-off (route 3), measure Z_{EQ}^{ME} yields no preference on where to locate the new RWC. The same holds for Z_{EQ}^{MAD}, which is due to the fact that this measure only considers the sum of the absolute deviations from the mean effect $\bar{c} = 0.5$, and disregards how these deviations are distributed over the different routes. One implication is that this score thereby provides no incentive to minimize the gaps among high coverage scores (i.e., above or equal to the average) and among low coverage scores (i.e., below or equal to the average), so that deviations 0.25 and 0.25 for respectively routes 1 and 2 are regarded as being equivalent to deviations 0.5 and 0. Measures Z_{EQ}^{SAD} and Z_{EQ}^{GC} avoid this by considering the entire distribution of the coverage scores. These measures provide an incentive to minimize the spread of these variables and thereby to avoid extreme values.

For the remainder of the chapter, we measure equity as Z_{EQ}^{SAD} and henceforth refer to this measure as the *equity score* Z_{EQ}. Note though that with some minor adaptations, the models and solution methods presented next can also be applied when equity is measured by Z_{EQ}^{MAD} or Z_{EQ}^{ME}.

4.4 Set Partitioning Type Formulation

De Vries et al. (2014b) propose a MIP formulation for the location problem defined in Section 4.3. This formulation uses binary variables x_k to indicate whether or not an RWC is placed at potential location k and i_{klq} to identify pairs of RWCs (k, l) that are adjacent along route q. The latter variables are used to calculate the travel

time gaps between facilities along the route (which yield the coverage scores c_q). We describe this formulation in Appendix 4.A and refer to it as the *direct formulation*.

This section proposes an alternative, set partitioning type of problem formulation, which we refer to as the *set partitioning formulation*. In contrast with the direct formulation, this formulation uses binary variables to indicate whether or not to establish a *configuration of RWCs* along a given route. A configuration of RWCs defines for each RWC location along a path whether or not an RWC is located there, as illustrated in Figure 4.3. The main advantage of introducing these decision variables is that this allows us to pre-calculate the coverage scores c_q, in contrast with the direct formulation, which defines c_q as a variable.

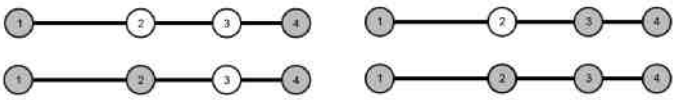


Figure 4.3: Four possible configurations of RWCs along truck route 1.

Next, we summarize the notations used in the set partitioning formulation:

Sets

KP: set of potential locations

Q: set of routes

N_q: set of possible configurations of RWCs along route q

N_q^k: subset of configurations of RWCs along route q with an RWC located at potential RWC location k

Decision variables

y_q^n: binary variable to indicate whether or not configuration $n \in N_q$ is chosen for route q

x_k: binary variable to indicate whether or not to locate an RWC at potential location k

$\Delta^+_{q_1 q_2}, \Delta^-_{q_1 q_2}$: continuous variables used to model the absolute difference between the coverage scores of routes q_1 and q_2.

Parameters

$\{w_{PV}, w_{CA}, w_{EQ}\}$: weights of the objective function terms

d_k: expected daily number of patients entering an RWC at potential location k

f_q: the number of truck drivers traveling route q

c_q^n: coverage score of route q given that configuration $n \in N_q$ is realized, as calculated by (4.4)

Using these notations, the location problem defined in Section 4.3 can be formulated as the following MIP model:

$$\max \quad w_{PV} \sum_{k \in KP} d_k x_k + w_{CA} \sum_{q \in Q} \sum_{n \in N_q} f_q c_q^n y_q^n$$
$$- w_{EQ} \sum_{q_1 \in Q} \sum_{\substack{q_2 \in Q \\ q_2 > q_1}} \left(\Delta^+_{q_1 q_2} + \Delta^-_{q_1 q_2} \right) f_{q_1} f_{q_2} \tag{4.9}$$

$$\text{s.t.} \sum_{n \in N_q} y_q^n = 1 \qquad q \in Q \tag{4.10}$$

$$\sum_{k \in KP} x_k = p \tag{4.11}$$

$$x_k - \sum_{n \in N_q^k} y_q^n = 0 \qquad k \in KP, q \in Q \tag{4.12}$$

$$\sum_{n \in N_{q_1}} c_{q_1}^n y_{q_1}^n - \sum_{n \in N_{q_2}} c_{q_2}^n y_{q_2}^n = \Delta^+_{q_1 q_2} - \Delta^-_{q_1 q_2} \qquad q_1, q_2 \in Q, q_2 > q_1 \tag{4.13}$$

$$x_k, y_q^n \in \{0, 1\} \qquad k \in KP, q \in Q, n \in N_q \tag{4.14}$$

$$\Delta^+_{q_1 q_2}, \Delta^-_{q_1 q_2} \geq 0 \qquad q_1, q_2 \in Q, q_2 > q_1 \tag{4.15}$$

Here, $\Delta^+_{q_1 q_2} = \max\{c_{q_1} - c_{q_2}, 0\}$ and $\Delta^+_{q_1 q_2} = \max\{-(c_{q_1} - c_{q_2}), 0\}$, so that $|c_{q_1} - c_{q_2}| = \Delta^+_{q_1 q_2} + \Delta^-_{q_1 q_2}$. The objective function (4.9) maximizes a weighted sum of the patient volume, the continuous access score and the equity score. Constraints (4.10) ensure that for each route $q \in Q$ only one configuration $n \in N_q$ is chosen. Next,

Constraint (4.11) ensures that the total number of newly placed RWCs is equal to p. The variables x_k and y_q^n are linked by Constraints (4.12), enforcing the choice of a configuration $n \in N_q^k$ if and only if x_k equals 1. Constraints (4.13) determine the value of $\Delta^+_{q_1 q_2}$ and $\Delta^-_{q_1 q_2}$. Finally, Constraints (4.14) define y_q^n and x_k as binary variables, and Constraints (4.15) define $\Delta^+_{q_1 q_2}, \Delta^-_{q_1 q_2}$ as non-negative variables. Note that this model is closely connected to the set partitioning problem, as Constraints (4.10) represent set partitioning constraints, and as we could express the model in terms of the "partitioning variables" y_q^n only (by substituting the variables x_k).

4.5 Column Generation Approach

Column generation has been proven to be a suitable technique for solving integer linear programming problems (ILPs) that involve a large number of variables. Instead of directly solving the linear programming relaxation of an ILP, called the Master Problem (MP), the method starts with a small subset of variables and solves the LP relaxation using this restricted variable set. This reduced problem is called the Restricted Master Problem (RMP). Using the resulting dual variables as an input, the method identifies excluded variables that may improve the solution value, and adds them to the set of included variables. The problem to find these variables is called the Pricing Problem. This process repeats till optimality is proven. Detailed information about column generation techniques can be found in Desaulniers et al. (2005) and Lübbecke (2010).

As our set partitioning formulation induces an exponential number of variables, we set out to develop a column generation approach to solve its LP relaxation. As we will show in Section 4.7, the set partitioning formulation has a very tight LP relaxation. Therefore, instead of embedding our column generation approach in a computationally expensive Branch-and-Bound scheme, we propose a heuristic that directly constructs a feasible integer solution from the LP relaxation solution. This section presents our basic solution approach, which solves the pricing problem exactly using a MIP formulation. Section 4.6 proposes several acceleration techniques, including an alternative method for solving the pricing problem.

4.5.1 Restricted Master Problem

The RMP is obtained by relaxing the variables x_k and y_q^n from the set partitioning formulation (i.e., by defining them as continuous variables instead of discrete variables) and by including only a subset of all variables y_q^n. We denote the subset of configurations n for route q for which the variable y_q^n is included in the RMP by N_q^*. Furthermore, we represent the dual variables associated with Constraints (4.10), (4.11), (4.12), and (4.13) by α_q, β, γ_{kq}, and $\delta_{q_1 q_2}$, respectively.

4.5.2 Pricing Problem

Generating a column for the RMP corresponds to finding a configuration $n \in N_q \setminus N_q^*$ for some route q for which the so-called reduced costs of variable y_q^n is positive. Let parameter a_{kq}^n indicate whether or not configuration n for route q has an RWC located at potential RWC location k. Then the reduced costs of the corresponding variable y_q^n are calculated as:

$$r_n^q = -\alpha_q + \sum_{k \in KP_q} a_{kq}^n \gamma_{kq} - c_q^n \rho_q \tag{4.16}$$

Here, $\rho_q = \sum_{u<q} \delta_{uq} - \sum_{u>q} \delta_{qu} - w_{CA} f_q$. Note that the reduced costs for y_q^n are independent of the characteristics of other routes, and hence that the pricing problem can be separated per route q:

$$\max_{n \in N_q} -\alpha_q + \sum_{k \in KP_q} a_{kq}^n \gamma_{kq} - c_q^n \rho_q \tag{4.17}$$

This problem can be regarded as an instance of the Roadside Healthcare Facility Location Problem with $|Q| = 1$, and hence can be solved using the MIP formulation presented in De Vries et al. (2014b). We describe the MIP formulation of (4.17) in Appendix 4.B.

4.5.3 Initialization and Termination Criterion

We start the column generation approach with a small subset of variables y_q^n for route q: the ones corresponding to configurations *without* new RWCs, configurations with *one* new RWC and configurations with an RWC at *each* potential RWC location along the route.

Solving the pricing problem for route q to optimality, we obtain one configuration n_q^* having reduced costs $r_q^* = r_{n_q^*}^q$. Because $y_q^n \leq 1$, LP duality implies:

$$z_{RMP} \leq z_{MP}^* \leq z_{RMP} + \sum_{q \in Q} r_q^* = UB$$

Here, z_{MP}^*, z_{RMP}, and UB denote the optimal solution value of the Master Problem, the optimal solution value of the RMP, and an upper bound on the optimal solution value of our location problem, respectively. The last equality follows from the observation that the MP is a relaxation of the location problem. Our column generation algorithm stops when $\frac{UB - z_{RMP}}{z_{RMP}} < \epsilon$, where ϵ is a fixed threshold.

4.5.4 Obtaining Integer Solutions

In each iteration (i.e., each time we solve the RMP), we use refined search to build a feasible solution from the fractional RMP solution. We first select the $p+s$ potential RWC locations with the largest value for x_k, where s denotes some strictly positive integer. Afterwards, we perform a refined search, identifying p out of the $p+s$ locations that yield a high solution value. Specifically, we first choose the $p-s$ locations with the largest value for x_k. Next, we complement these by s out of the $2s$ remaining locations, evaluating each of the $\binom{2s}{s}$ possibilities. If the resulting solution has a higher solution value then the best known integer solution, we replace the best solution by the current solution.

Our numerical experiments show that such heuristic solution often has a higher solution value than the corresponding solution of the RMP. We therefore add the variables corresponding to the configurations that produce a heuristic solution to the variable set of the RMP.

4.6 Acceleration Strategies

Column generation solution approaches often converge slowly towards the optimum. Causes include a large number of iterations needed to reach convergence, and large solution times for the pricing problem and/or the restricted master problem. This section proposes five acceleration strategies tackling these causes: accelerated pricing, adding multiple columns, column pool management, dual stabilization, and a 2-stage approach.

4.6.1 Accelerated Pricing

Solving the pricing problem using the MIP model is computationally expensive. Instead of directly solving the pricing problem to optimality, the method presented next tries to quickly identify an attractive (note, not necessarily optimal) variable using a sequence of shortest path problems, and only uses the MIP model if no attractive column is identified. Next, we introduce some theoretical results inspiring this approach.

Let e_q^n represent the fraction of time a truck driver traveling route q is "safe" when choosing configuration n (see Section 4.3.2). Furthermore, let $g_1(e_q^n) = 0$, $g_2(e_q^n) = \frac{e_q^n - \mu_1}{\mu_2 - \mu_1}$, and let $g_3(e_q^n) = 1$, i.e. these functions are the extrapolations of the three function segments of $g(\cdot)$. Next, consider the set of configurations n for which e_q^n lies in a given segment σ_i: N_{qi}. We relate to this set the acyclic graph G_{qi} depicted in Figure 4.4. This graph has one node for each RWC location along route q, including origin node $O := KC_q(1)$ and destination node $D := KC_q(m)$. Let ϕ^n denote the $O-D$ path visiting the nodes corresponding to the RWCs in configuration n. In Appendix 4.C, we prove the following result:

Proposition 4.1. *There exist arc weights for graph G_{qi} such that for each $n \in N_{qi}$ holds that the length of the corresponding $O - D$ path $l(\phi^n)$ equals $-r_q^n$*

Proof. See Appendix 4.C □

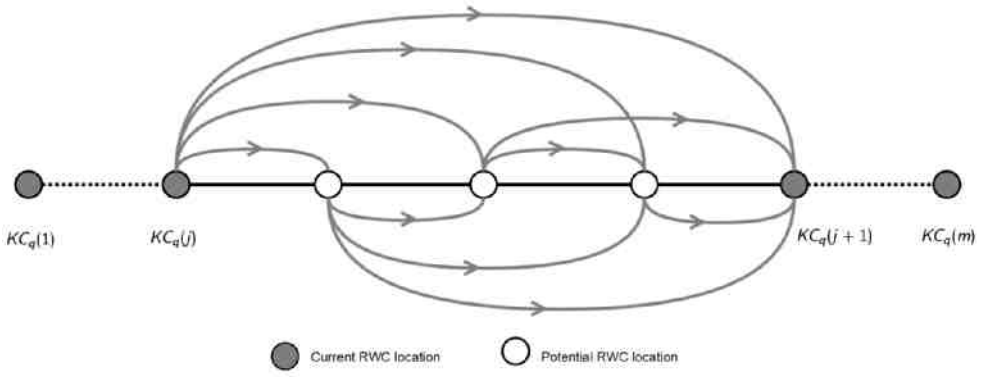


Figure 4.4: Directed graph G_{qi} for route q. Each node corresponds to an RWC location. The nodes visited in a path ϕ^n from $KC_q(1)$ to $KC_q(m)$ correspond to the RWCs located in configuration n.

A direct implication of Proposition 4.1 is that configurations $n \in N_{qi}$ for which r_q^n is larger correspond to shorter paths in the acyclic graph. This inspires the following solution approach for the pricing problem corresponding to route q (see Algorithm 1). We construct the acyclic graph corresponding to segment σ_1 and solve the shortest path problem for this graph using the Bellman-Ford algorithm. Next, we calculate $r_q^{\hat{n}}$ for the configuration $\hat{n}$ corresponding to the shortest path. If $r_q^{\hat{n}} > 0$, we mark variable $y_q^{\hat{n}}$ as an attractive variable. We repeat this procedure for segment σ_2, and add the attractive variable having the largest reduced costs to the set of variables included in the RMP. (The proof of Proposition 4.1 implies that the configurations obtained for segments σ_1 and σ_3 are the same, so that we do not need to repeat the procedure for σ_3).

Note that this first part of our solution approach is not exact. Since the graph corresponding to segment σ_i misrepresents the reduced costs corresponding to configurations $n \notin N_{qi}$, the configuration found may not be the one with the largest reduced costs. In some cases, however, we can use the configurations obtained to prove that no configuration with positive reduced costs exists. For example, let the configurations found for the graphs G_{q1} and G_{q2} be represented by n_1 and n_2, respectively. Suppose that $e_q^{n_1} \in \sigma_3$ and $e_q^{n_2} \in \sigma_2$, that $0 \geq r_q^{n_2} \geq r_q^{n_1}$, and that $\rho_q > 0$.

By optimality of n_2 know that for each of the configurations n for which $e_q^n \in \sigma_2$ holds that $r_q^n \leq r_q^{n_2}$. Furthermore, because $\rho_q > 0$, we know that for each of the configurations n for which $e_q^n \in \sigma_1 \cup \sigma_3$ holds that $r_q^n \leq r_q^{n_1} \leq r_q^{n_2}$ (see Equation (4.16)), showing that for each n holds that $r_q^n \leq r_q^{n_2} \leq 0$.

In case that we cannot prove that a column with positive reduced costs does not exist, we set out to find one using a refined search. Specifically, we propose to solve the pricing problem $|KP_q|$ times. In each time, we adapt the graphs G_{qi} such that they enforce locating an RWC at a specific potential RWC location $k \in KP_q$ and optimize the other location decisions using the Bellman-Ford algorithm. Next, we add the attractive variable having the largest reduced costs to the set of variables included in the RMP. If also the refined search fails to identify an attractive variable, we solve the pricing problem using the MIP model presented in Appendix 4.B.

Algorithm 1 Pricing algorithm for route q.

Require:
route q and dual variables α_q, β, γ_{kq}, and $\delta_{q_1 q_2}$
Ensure:
variable with positive reduced cost or certificate that no such variable exists

construct graphs G_{q1} and G_{q2}
obtain configuration(s) using the Bellman-Ford algorithm
if reduced costs of at least one of the corresponding variables are positive **then**
    return the variable with the largest reduced costs
else if certificate that no such variable exists **then**
    stop
else apply refined search
    if variable with positive reduced cost found **then**
        return the variable with the largest reduced costs
    else
        solve pricing problem as a MIP
        return a variable with positive reduced costs or a certificate that no such variable exists
    end if
end if

Note that we apply this solution method to each route q, yielding up to $|Q|$ variables to be included in the RMP per iteration.

4.6.2 Adding Multiple Columns

The algorithm presented in the previous subsection adds in each iteration at most one variable per route to the set of variables included in the RMP. Adding *multiple* attractive variables per iteration is a common acceleration technique. The ideas underlying it are that any variable with positive reduced costs is attractive for entering the RMP, that it is often easy to find multiple attractive variables, and that adding multiple variables tends to significantly decrease the number of iterations required to solve the RMP. We propose to adapt steps 3 - 7 of Algorithm 1 as follows: we start with obtaining the optimal configuration(s) corresponding to the two generated graphs *and* perform the refined search algorithm. All variables with positive reduced costs found are then added to the variable set of the RMP.

4.6.3 Column Pool Management

We save each column generated in our the column generation approach in a column pool. Before each iteration, we determine for each column in the pool whether or not to include it in the RMP. The aim is to balance the purpose to maximize the objective value of the RMP (by including as many attractive columns as possible) and on the other hand the aim to minimize the solution time for the RMP (by minimizing the number of columns included). Specifically, we include a column if at least one of the following conditions holds: (1) it was generated in the initialization, (2) it resulted from the previous call to the pricing problem, (3) it has been part of the basis at least once in the last κ iterations, where κ denotes some strictly positive integer (4) it has reduced costs greater than a given percentage of the best reduced cost among all columns available for that path, (5) it is part of an integer solution found that has an objective value greater than the corresponding RMP relaxation.

4.6.4 Dual Stabilization

Column generation is often characterized by large oscillations in the dual variables from iteration to iteration. This leads to slow convergence and often to degeneracy. Several techniques have appeared that stabilize the dual variables and thereby accelerate the convergence (see Desaulniers et al., 2005). We propose to call the pricing problem twice per iteration. The first call provides the dual solution obtained in the current iteration, π^*, as an input. The second call provides $\pi_{ST} = \lambda\hat{\pi} + (1-\lambda)\pi^*$ as an input, as suggested by Pessoa et al. (2010). Here, $\hat{\pi}$ represents the best known dual solution (i.e., the dual that produces the smallest upper bound) and λ denotes some constant between 0 and 1.

4.6.5 2-Stage Approach

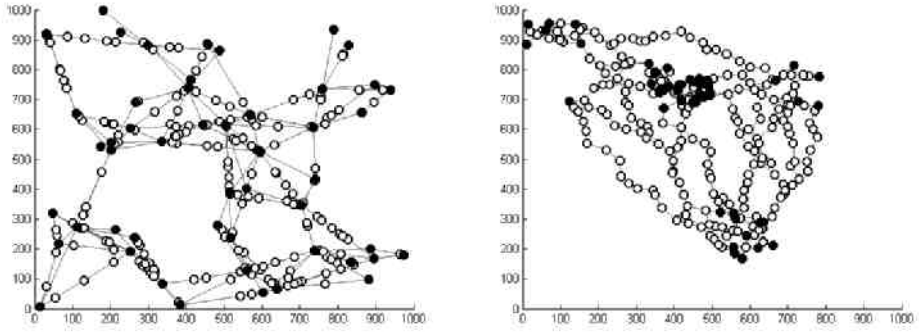
The time required to solve the RMP is relatively large. This seems to be caused by the large number of variables and constraints needed to calculate the equity score. In response to that, we implement a 2-Stage column generation approach. In the 1^{st} stage we solve the RMP without Constraints (4.13). This corresponds to leaving the equity score out of consideration. In the 2^{nd} stage, we include Constraints (4.13), and solve the full RMP with the columns generated so far.

4.7 Computational Experiments

This section numerically analyzes the column generation approach presented in Sections 4.5 and 4.6. Section 4.7.1 starts with describing the randomly generated networks used for this analysis. Next, we analyze the integrality gap for the direct formulation and the set partitioning formulation in Section 4.7.2. For our baseline case, Section 4.7.3 describes the convergence of the column generation approach and the impact of each of the acceleration strategies. Finally, Section 4.7.4 analyzes the computational performance of the column generation approach in detail for the entire set of instances. All mathematical models are implemented in Java, using ILOG CPLEX v12.61 to solve our LP and MIP problems, and analyzed using a computer running Windows 8.1 with a 2.3 GHz i7-4712HQ processor and 16GB of RAM.

4.7.1 Randomly Generated Instances

We generate clustered and non-clustered network graphs to test our algorithm. Appendix 4.D describes the methods we use to generate the graphs, routes, and demand parameters in detail. Furthermore, Figures 4.5a and 4.5b show an example of a non-clustered and a clustered instances with 200 potential locations and 200 routes.



(a) Non-clustered instance. **(b)** Clustered instance.

Figure 4.5: Examples of randomly generated instances. Black points represent current RWC locations and white points represent potential RWC locations.

	Instance Names	Quantity	n_{OD}	n_r	n_p
Medium	r75p150n1, r75p150n2,...	10	15	5	150
	r125p150n1, r75p150n2,...	10	25	5	150
	r200p100n1, r200p100n2,...	10	40	5	100
	r200p200n1, r200p200n2,...	10	40	5	200
	r200p200c1, r200p200c2,...	10	40	5	200
Large	r320p600n1, r320p600n2,...	3	80	4	600
	r500p1000n1, r500p1000n2,...	3	100	5	1000
	r2400p5000n1, r2400p5000n2,...	3	300	8	5000

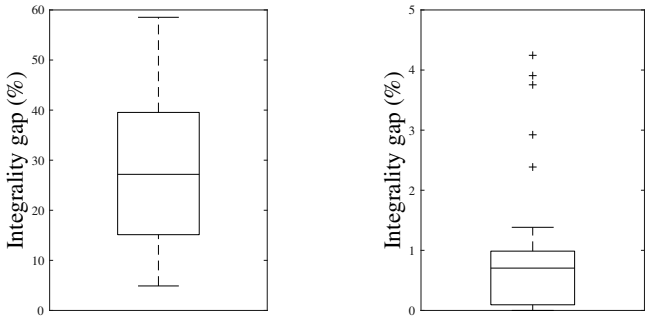
Table 4.3: Characteristics of the randomly generated instances.

Table 4.3 describes the characteristics of 59 randomly generated instances used for our numerical experiments. Here, n_{OD} denotes the number of O-D nodes in the generated graphs, n_r represents the number of routes departing from each O-D node

(hence, the total number of routes equals $n_r \cdot n_{OD}$), and n_p denotes the number of potential RWC locations in the generated instance. An instance named rXpYnZ resp. rXpYcZ is the Z^{th} instance in the set of non-clustered resp. clustered instances having X routes and Y potential locations. Furthermore, we refer to instance $r200p200n1$ as our baseline case. Finally, we use the following parameter values in our experiments: $p = 20, \mu_1 = 0.4, \mu_2 = 0.8, \tau = 100, w_{PV} = 10, w_{CA} = 1.5$ and $w_{EQ} = 10^{-4}$. The values of the latter three parameters were chosen so as to reflect the trade-offs decision makers face in practice. The instances can be accessed at `http://people.few.eur.nl/hdevries`.

4.7.2 Integrality Gap

Figure 4.6 shows boxplots of the integrality gap (i.e., the gap between the optimal solution value and the optimal solution value of the LP relaxation) for the direct formulation and the set partitioning formulations, based on the 50 medium instances (large instances were excluded, as many could not be solved to optimality). We observe that the integrality gap for the direct formulation is much larger: it generally ranges between 20% and 40% whereas the gap for the set partitioning formulation is generally smaller than 1%. Furthermore, solving the LP relaxation of the direct formulation results in an integer (i.e., an optimal) solution for 20% of the non-clustered instances, highlighting the strength of the formulation.



(a) Direct formulation. **(b)** Set part. formulation.

Figure 4.6: Integrality Gap.

4.7.3 Effect of Acceleration Strategies

Figure 4.7 shows the convergence of the basic column generation approach for our baseline case $r200p200n1$. It confirms the slow convergence behavior that has motivated several of our acceleration techniques. The effects of these techniques in terms of the optimality gap reached after 500 seconds (note, we know the optimal solution by solving the location problem exactly), the time/number of iterations needed to reach convergence, the time/number of iterations needed to find the final solution (i.e., the best solution found), and the total number of columns generated/included in the last iteration are described for the baseline case in Table 4.4. This table provides these statistics for the following solution approaches: (1) the basic approach (BA), as presented in Section 4.5, (2) approach 1 + accelerated pricing (AP), (3) approach 2 + adding multiple columns (MC), (4) approach 3 + column pool management (CM), (5) approach 4 + dual stabilization (DS), and (6) approach 5 + 2-Stage approach (2S).

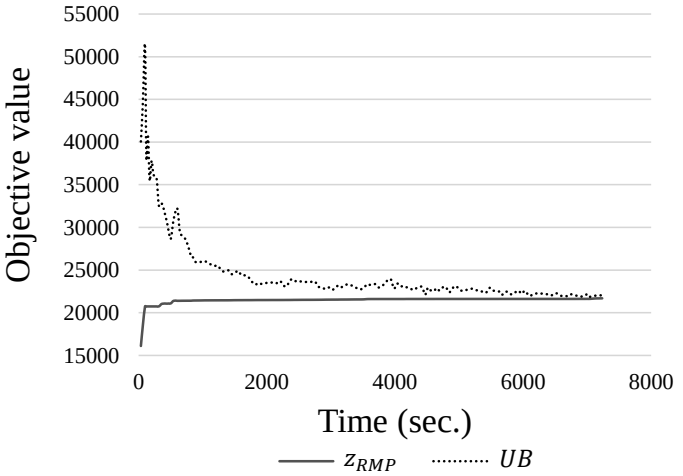


Figure 4.7: Convergence of the basic column generation approach for instance r200p200n1.

Solution approach	**1**	**2**	**3**	**4**	**5**	**6**
Opt. gap after 500 sec. (%)	2.314	0.759	1.130	0.000	0.684	0.000
Time final solution (sec.)	3026	6368	1410	495	946	375
Time convergence (sec.)	>7200	>7200	1410	1387	946	393
Iterations to final solution	57	120	28	10	22	23
Iterations to convergence	>130	>133	28	35	22	24
Columns last iteration	7573	5764	14432	13232	12616	9073
Total columns generated	7573	5764	14432	16706	18458	10031

Table 4.4: Impact of acceleration strategies for instance $r200p200n1$. Solution approaches: 1 = BA, 2 = 1 + AP, 3 = 2 + MC, 4 = 3 + CM, 5 = 4 + DS, and 6 = 5 + 2S.

Though each approach identified the optimal solution, the table shows that the acceleration techniques jointly result in a vast improvement in the computational performance of the column generation algorithm. Particularly adding multiple columns, column pool management, and the 2-stage approach have a large impact. This seems to be explained by the fact that MC greatly reduces the number of iterations and that CM and 2S significantly reduce the time required to solve the RMP (which accounts for approximately 99% of the solution time).

Figure 4.8 provides a more detailed view on the performance of the acceleration techniques, showing the development of the solution value for solution approaches 1, 4, and 6 (for sake of clarity we excluded the others from the figure). It shows that the approach including all acceleration techniques obtains near-optimal solutions within a couple of seconds.

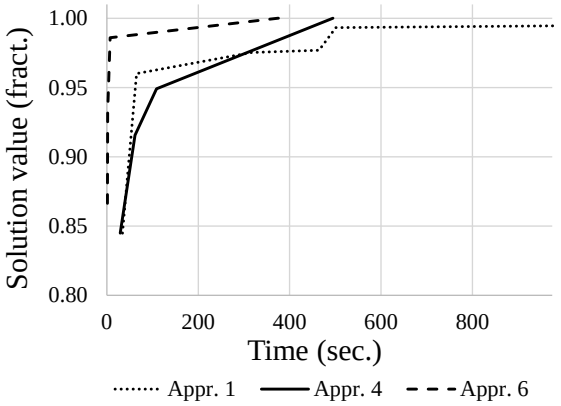


Figure 4.8: Development of the solution value over time for solution approaches 1 (BA), 4 (BA + AP + MC + CM), and 6 (BA + AP + MC + CM + DS + 2S), measured as a fraction of the optimal solution value.

4.7.4 Overall Results

Next, we analyze the performance of the column generation approach for the full set of instances and compare it with the performance of CPLEX 12.61 on the direct formulation. The specific column generation approach we consider here includes each of the acceleration strategies, and will be referred to as the 2-stage approach from now on. Furthermore, we use a computation time of 1 hour as a stopping criterion.

Table 4.5 describes the performance of the two approaches for the 50 medium instances. Columns t_{CONV} and t_{SOL} contain the average time to convergence and the average time to finding the final solution (i.e., the best solution found) for both approaches and column *# Converged* shows the number of instances for which conver-

gence was reached within 1 hour of computation time. They show that average solution times for both approaches are comparable and that the approach using CPLEX on the direct formulation seems to outperform the 2-stage approach in terms of time to convergence and in terms of the number of times convergence was reached within the time limit. The 2-stage approach obtains the optimal solution in 34 of the 50 instances (all of them were non-clustered instances), and yields a negligible optimality gap for the others. Columns t_{RMP} and t_{PP}, providing the average time this approach spends on the RMP and the pricing problem, highlight the efficiency of our approach for solving the pricing problem and stress the importance to decrease computation times for the RMP. Finally, we note that the performance of both approaches is worse for the clustered instances. The reason is that the clustering induces several routes that pass many potential RWC locations. As the number of variables and constraints in the direct formulation and the number of variables in the set partitioning formulation increases steeply with the number of potential facility locations along a path (see Table 4.6), this substantially increases their complexity.

	Avg. opt. gap		# Converged		Avg. computation time					
					DF		2S			
	DF	2S	DF	2S	t_{SOL}	t_{CONV}	t_{SOL}	t_{CONV}	t_{RMP}	t_{PP}
r75p150n	0.00%	0.01%	10	10	3.28	4.75	5.61	12.05	3.32	0.30
r125p150n	0.00%	0.03%	10	10	44.38	88.84	19.75	143.52	73.70	5.09
r200p100n	0.00%	0.03%	10	10	30.11	57.49	40.45	60.28	32.29	0.53
r200p200n	0.00%	0.00%	10	9	133.84	230.05	169.2	1056.68	890.1	7.05
r200p200c	0.24%	0.87%	8	6	903.12	1468.53	1128.57	2396.76	2029.11	6.19

Table 4.5: Performance of CPLEX 12.61 on the direct formulation (DF) and the column generation approach including each of the acceleration strategies (2S) on the 50 medium problem instances.

Table 4.6 describes some more details about the 2-stage solution approach, including the average number of iterations till convergence, the average number of columns included in the last iteration, the average total number of columns generated during the execution of the algorithm, and the average total number of columns in the instances (i.e., included in the master problem). The huge number of columns corresponding to the larger instances show that a column generation solution approach is crucial for the set partitioning formulation.

	#Iterations	#Columns		
		Last Iteration	Generated	Total
r75p150n	32	4,224	4,950	345,623
r125p150n	60	9,704	16,871	976,515
r200p100n	23	3,035	3,158	19,237
r200p200n	58	14,677	32,164	2,476,941
r200p200c	85	20,265	61,396	423,362,894

Table 4.6: Average numbers of iterations performed and columns generated by the 2-stage approach on the 50 medium problem instances.

The computational benefits of the 2-stage approach become particularly visible when large problem instances are considered. Though both formulations did not converge after the time limit of 1 hour, Table 4.7 shows that the value of the best solution obtained by the 2-stage approach is on average much higher. For the set of largest instances, CPLEX even fails to produce an integer solution within the time limit.

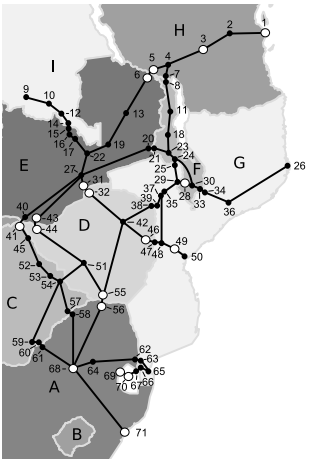
Instance	Obj. value		Difference
	DF	2S	
r320p600n1	16888.75	17001.31	112.57
r320p600n2	24863.54	24704.83	-158.70
r320p600n3	23539.43	22933.96	-605.48
r500p1000n1	13397.37	14759.82	1362.45
r500p1000n2	7977.70	8594.52	616.81
r500p1000n3	-6661.87	15562.55	22224.42
r2400p5000n1	-	227090.11	-
r2400p5000n2	-	222021.25	-
r2400p5000n3	-	222671.04	-

Table 4.7: Objective value obtained after 1 hour for the 9 large problem instances using approaches DF and 2S.

4.8 Sensitivity Analysis on Equity

As motivated, it is becoming more and more relevant to have insight in the trade-off between the equity criterion and North Star's current optimization criteria. This section investigates this trade-off for the case study introduced by De Vries et al. (2014b). This case considers the North South Corridor Network, a network of high-

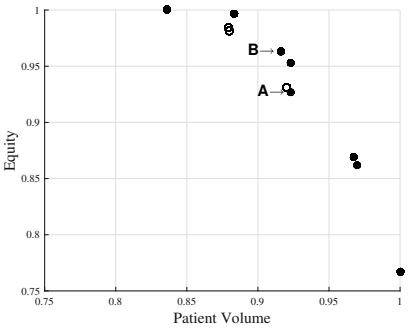
ways connecting the main regions of economic importance in Southern and East Africa with the main regional ports. The case includes 53 potential RWC locations, 18 current RWC locations, and 60 long distance truck routes. Figure 4.9 depicts this network and the locations of current RWCs in this network (note, the RWC Equivalents at truck route origins and destinations are not depicted here). The case study data can be accessed at `http://people.few.eur.nl/hdevries`.



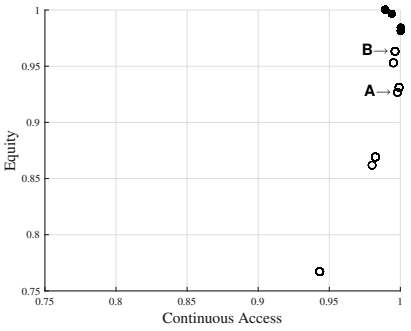
1	Dar es Salaam Port	22	Kapiri Mposhi	43	Victoria Falls	64	Middelburg
2	Morogoro	23	Lilongwe	44	Livingstone	65	Maputo Port
3	Iringa	24	Dedza	45	Pandamatenga	66	Lomahasha
4	Uyole	25	Calomue	46	Forbes	67	Namaacha
5	Tunduma	26	Nacala Port	47	Machipanda	68	Johannesburg
6	Nakonde	27	Lusaka	48	Chimoio	69	Mbabane
7	Kasumulu	28	Mwanza	49	Inchope	70	Manzini
8	Songwe	29	Zobue	50	Beira Port	71	Durban Port
9	Kolwesi	30	Blantyre	51	Bulawayo		
10	Likasi	31	Chirundu North	52	Nata	A	South Africa
11	Mzuzu	32	Chirundu South	53	Sebina	B	Lesotho
12	Lubumbashi	33	Muloza	54	Francistown	C	Botswana
13	Mpika	34	Milange	55	Beitbridge North	D	Zimbabwe
14	Kasumbalesa North	35	Tete	56	Beitbridge South	E	Zambia
15	Kasumbalesa South	36	Mocuba	57	Martins Drift	F	Malawi
16	Chililabombwe	37	Changara	58	Goblersbrug	G	Mozambique
17	Kitwe	38	Nyamapanda	59	Gaborone	H	Tanzania
18	Kasungu	39	Cuchamano	60	Tlokweng	I	D.R. Congo
19	Serenje	40	Kazungula North	61	Kopfontein		
20	Mwami	41	Kazungula South	62	Lubombo	•	Potential RWC
21	Mchinji	42	Harare	63	Ressano Garcia	○	Current RWC

Figure 4.9: Map of the North-South Corridor Network.

We solve the problem to place 5 new RWCs in this network for 1000 settings of the weight parameters w_{PV}, w_{CA}, and w_{EQ}, yielding 11 Pareto efficient solutions. Figures 4.10a and 4.10b show their relative optimality with respect to each of the three criteria, measured as a fraction of the highest attainable value for these criteria. The filled dots in such figure represent solutions that are not only Pareto efficient with respect to the three criteria, but also with respect to the two criteria considered in the figure.



(a) Quality with respect to patient volume and equity.



(b) Quality with respect to continuous access and equity.

Figure 4.10: Quality of 11 Pareto-efficient solutions with respect to two optimization criteria, measured as a fraction of the highest attainable value for these criteria (based on 5 new facilities).

We see that considerable gains in terms of equity can be made at a marginal cost in terms of continuous access and/or patient volume. For example, choosing the solution corresponding to point B (equity is assigned a significant weight) in these figures instead of the solution corresponding to point A (equity is hardly considered) increases the equity score by 3.6 percentage points and decreases the continuous access score by only 0.1 percentage points and the patient volume by only 0.7 percentage points. More generally, our results suggest that it is possible to obtain solutions that are relatively close to optimal with respect to each optimization criterion. For example the relative optimality of the solution corresponding to point A with respect to *each* of the three criteria equals at least 92%.

Figure 4.11 illustrates the impact of including the equity criterion on the actual levels of access provided in solutions A and B by showing the distribution of the coverage scores for the entire truck driver population we consider. Increasing the importance of equity decreases the number of truck drivers having very low coverage levels or high coverage levels. For example, the number of truck drivers having very low coverage levels decreases from 370 to 0 and the number of truck drivers having high coverage levels decreases by 21.

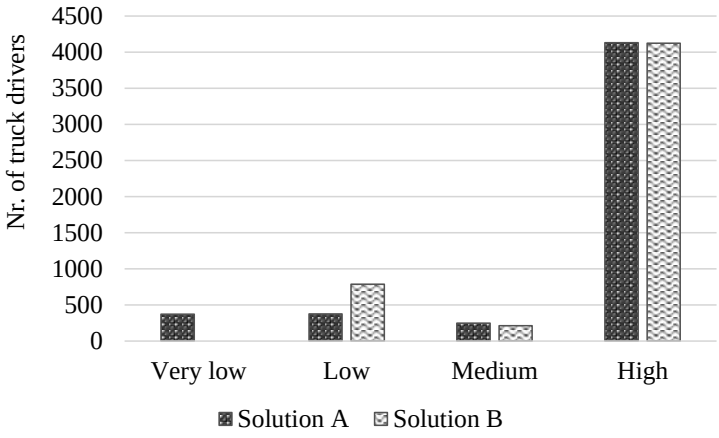


Figure 4.11: Coverage scores for truck driver population in Pareto efficient solutions A (equity is hardly considered) and B (equity is assigned a significant weight). Very low: $c_q \in [0, 0.25)$, low: $c_q \in [0.25, 0.5)$, medium: $c_q \in [0.5, 0.75)$, high: $c_q \in [0.75, 1]$.

4.9 Conclusions and Discussion

This chapter considers the problem to locate a given number of new RWCs based on three optimization criteria: maximize patient volume, ensure continuity of access, and establish an equitable health system. We first propose several measures of equity of access to healthcare in the context of a mobile patient population, based on a review of the literature available. Measures for the patient volume criterion and the continuous access criterion are taken from previous work on this problem. The equity criterion significantly increases the complexity of the location problem, and existing models and solution methods will be unsuitable for solving large instances. We therefore come up with a novel, set-partitioning type of formulation for the problem. Numerical experiments show that, in contrast with the formulation previously introduced, the integrality gap for this formulation is very small. As this formulation requires an exponential number of variables, we propose a column generation algorithm to solve it. Additionally, we propose and analyze several state-of-the-art

acceleration techniques, including accelerated pricing, dual stabilization and a 2-stage approach.

Though the facility location problem is strongly $\mathcal{NP}$-hard, our algorithm yields near-optimal solutions to large problem instances within an acceptable amount of time. Nevertheless, we see several opportunities for improvement. For example, our approach can be transformed into an exact approach by embedding it in a Branch-and-Price scheme. Furthermore, we use a relatively simple heuristic for generating feasible integer solutions. More advanced (meta)heuristics, taking better advantage of the benefits of placing multiple facilities along a route, could to be promising.

Finally, our numerical analysis of the trade-off between equity and North Star's current optimization criteria provides some interesting insights. Considerable gains in terms of equity can be made at a marginal loss in terms of patient volume and/or continuous access. Furthermore, our results suggest that solutions that are close to optimal in terms of each of the three criteria are attainable by appropriately balancing the weights attached to them. This provides a strong argument for incorporating the equity criterion into future network design decisions.

Appendices

4.A Direct Formulation

The direct formulation uses the following notations in addition to those introduced earlier in this chapter. K_{kq} represents the set of locations that are passed *after* location k during a trip from along route q. Variable $i_{klq}, k \in K_q, l \in K_{kq}$ equals 1 if locations k and l are adjacent RWCs along route q (i.e., RWCs are located at both locations, and there is no RWC between them), and equals 0 otherwise. Finally, λ_{iq} and z_{iq} are auxiliary variables to model the piecewise linear function $g(\cdot)$

$$\max \quad w_{PV} \sum_{k \in K} d_k x_k + w_{CA} \sum_{q \in Q} f_q c_q$$

$$- w_{EQ} \sum_{q_1 \in Q} \sum_{\substack{q_2 \in Q \\ q_2 > q_1}} \left(\Delta^+_{q_1 q_2} - \Delta^-_{q_1 q_2} \right) f_{q_1} f_{q_2} \tag{4.18}$$

$$\text{s.t. } c_q = \lambda_{3q} + \lambda_{4q} \qquad q \in Q \tag{4.19}$$

$$\lambda_{1q} 0 + \lambda_{2q} \mu_1 + \lambda_{3q} \mu_2 + \lambda_{4q} = \frac{1}{T_q} \left(\sum_{k \in K_q} \sum_{l \in K_{kq}} i_{klq} \min\{t_{kl}, \tau\} \right) q \in Q \tag{4.20}$$

$$\lambda_{1q} + \lambda_{2q} + \lambda_{3q} + \lambda_{4q} = 1 \qquad q \in Q \tag{4.21}$$

$$\lambda_{1q} \leq z_{1q} \qquad q \in Q \tag{4.22}$$

$$\lambda_{2q} \leq z_{1q} + z_{2q} \qquad q \in Q \tag{4.23}$$

$$\lambda_{3q} \leq z_{2q} + z_{3q} \qquad q \in Q \tag{4.24}$$

$$\lambda_{4q} \leq z_{3q} \qquad q \in Q \tag{4.25}$$

$$z_{1q} + z_{2q} + z_{3q} = 1 \qquad q \in Q \tag{4.26}$$

$$c_{q_1} - c_{q_2} = \Delta^+_{q_1 q_2} - \Delta^-_{q_1 q_2} \qquad q_1, q_2 \in Q, q_2 > q_1 \tag{4.27}$$

$$x_k = 1 \qquad k \in KC \tag{4.28}$$

$$\sum_{k \in KP} x_k = p \tag{4.29}$$

$$\sum_{l \in K_q} i_{klq} = x_k \qquad k \in K_q, q \in Q \tag{4.30}$$

$$\sum_{l \in K_q} i_{klq} = 1 \qquad k \in KC_q(1), q \in Q \tag{4.31}$$

$$\sum_{k \in K_q} i_{klq} = x_l \qquad l \in K_q, q \in Q \tag{4.32}$$

$$\sum_{k \in K_q} i_{klq} = 1 \qquad l \in KC_q(m), q \in Q \tag{4.33}$$

$$x_k \in \{0, 1\} \qquad k \in K_q, q \in Q \tag{4.34}$$

$$\Delta^+_{q_1 q_2}, \Delta^-_{q_1 q_2} \geq 0 \qquad q_1, q_2 \in Q, q_2 > q_1 \tag{4.35}$$

$$i_{klq} \in [0, 1] \qquad k \in K_q, l \in K_{kq}, q \in Q \tag{4.36}$$

$$\lambda_{iq} \geq 0 \qquad i \in \{1, 2, 3, 4\}, q \in Q \tag{4.37}$$

$$z_{iq} \in \{0, 1\} \qquad i \in \{1, 2, 3\}, q \in Q \tag{4.38}$$

The objective function (4.18) represents a weighted sum of the patient volume, the continuous access score, and the equity score. Constraints (4.20) - (4.26) define c_q

as a piece-wise linear function of the fraction of time a truck driver traveling route q is "safe" during his trip. This fraction is calculated by the right-hand side of (4.20). Next, Constraints (4.27) enable the calculation of $|c_{q_1} - c_{q_2}|$. The current network of RWCs is described in Constraint (4.28). Constraints (4.30)-(4.33) ensure that the variables i_{klq} attain the correct value. Finally, the decision variables are defined in (4.34) - (4.38).

4.B MIP Formulation Pricing Problem

$$\max \quad -\alpha_q + \sum_{k \in KP_q} \gamma_{kq} x_k - \rho_q c_q \tag{4.39}$$

$$\text{s.t. } c_q = \lambda_{3q} + \lambda_{4q} \tag{4.40}$$

$$\lambda_{1q} 0 + \lambda_{2q}\mu_1 + \lambda_{3q}\mu_2 + \lambda_{4q} = \frac{1}{T_q}\left(\sum_{k \in K_q} \sum_{l \in K_{kq}} i_{klq} \min\{t_{kl}, \tau\}\right) \tag{4.41}$$

$$\lambda_{1q} + \lambda_{2q} + \lambda_{3q} + \lambda_{4q} = 1 \tag{4.42}$$

$$\lambda_{1q} \leq z_{1q} \tag{4.43}$$

$$\lambda_{2q} \leq z_{1q} + z_{2q} \tag{4.44}$$

$$\lambda_{3q} \leq z_{2q} + z_{3q} \tag{4.45}$$

$$\lambda_{4q} \leq z_{3q} \tag{4.46}$$

$$z_{1q} + z_{2q} + z_{3q} = 1 \tag{4.47}$$

$$x_k = 1 \qquad k \in KC_q \tag{4.48}$$

$$\sum_{l \in K_q} i_{klq} = x_k \qquad k \in K_q \tag{4.49}$$

$$\sum_{l \in K_q} i_{klq} = 1 \qquad k \in KC_q(1) \tag{4.50}$$

$$\sum_{k \in K_q} i_{klq} = x_l \qquad l \in K_q \tag{4.51}$$

$$\sum_{k \in K_q} i_{klq} = 1 \qquad l \in KC_q(m) \tag{4.52}$$

$$x_k \in \{0,1\} \qquad k \in K_q \tag{4.53}$$

$$i_{klq} \in [0,1] \qquad k \in K_q, l \in K_{kq} \tag{4.54}$$

$$\lambda_{iq} \geq 0 \qquad i \in \{1,2,3,4\} \tag{4.55}$$

$$z_{iq} \in \{0,1\} \qquad i \in \{1,2,3\} \tag{4.56}$$

See Appendix 4.A for the interpretation of this model. Solving the problem formulated in this model yields location decisions x_k, which define a configuration of RWCs n for route q.

4.C Proof of Proposition 4.1

Let K_q^n denote the ordered set of RWC locations along route q for which $x_k = 1$ in configuration n. Our proof of Proposition 4.1 makes use of the following lemma:

Lemma 4.1. *Suppose that e_q^n lies in segment σ_i of the piecewise linear function $g(\cdot)$. Then*

$$c_q^n = \sum_{j=1}^{m-1} g_i \left(\frac{\min\{t_{\left(K_q^n(j), K_q^n(j+1)\right)}, \tau\}}{T_q} \right) \tag{4.57}$$

Proof. This immediately follows from definition (4.4), and from the fact that the linearity of $g_i(\cdot)$ implies that the summation can be taken out of this function. □

Proposition. There exist arc weights for graph G_{qi} such that for each $n \in N_{qi}$ holds that the length of the corresponding $O-D$ path $l\left(\phi^n\right)$ equals $-r_q^n$.

Proof. We define the weights corresponding to the arcs in this graph as follows. First, we add α_q to all arcs departing from the $K_q(1)$ (the current RWC location at the origin of the route) and add $-\gamma_{kq}$ to all incoming arcs for potential RWC location $k \in KP_q$. Finally, we add $g_i\left(\frac{\min\{t_{kl},\tau\}}{T_q}\right)\rho_q$ to each arc (k,l).

To prove our claim, let us define $l\left(\phi^n\right)$. Note that path ϕ^n must use an arc departing from the node corresponding to $K_q(1)$, so that it's length includes the

term α_q. Furthermore, it includes the term $-\gamma_{kq}$ for each visited node corresponding to a potential RWC location $k \in KP_q$. Finally, the length includes the term $g_i\left(\frac{\min\left\{t_{\left(K_q^n(j),K_q^n(j+1)\right)},\tau\right\}}{T_q}\right)\rho_q$ for each of the arcs corresponding to the pair of RWC locations $\left(K_q^n(j), K_q^n(j+1)\right)$. Hence, using Lemma 4.1 we obtain that:

$$\begin{aligned} l\left(\phi^n\right) &= \alpha_q - \sum_{k\in KP_q} a_{kq}^n \gamma_{kq} + \sum_{j=1}^{m-1} g_i\left(\frac{\min\{t_{\left(K_q^n(j),K_q^n(j+1)\right)},\tau\}}{T_q}\right)\rho_q \\ &= \alpha_q - \sum_{k\in KP_q} a_{kq}^n \gamma_{kq} + c_q^n \rho_q = -r_q^n \end{aligned}$$

This concludes the proof. □

4.D Random Network Generation

The non-clustered instances were generated as follows. First, we generate the locations of n_{OD} O-D nodes in $[0, 1000]^2$ according to a continuous uniform distribution. We calculate Euclidian distances between nodes, and take the minimum spanning tree of the full graph. Next, the full road network is obtained by adding for each node n_a additional arcs, connecting it to the n_a closest nodes it has not yet been connected to in the minimum spanning tree. We assume that each O-D node represents a current RWC location (in accordance with Assumption 4.1), and that currently no RWC is located elsewhere. The locations of n_p potential RWC locations are generated by drawing completely random positions at the full road network (hence, each location along the full road network has equal probability of being selected). Next, we generate for each O-D node n_r truck routes that have this node as an origin, and randomly select one of the other O-D nodes as its destination (if a generated flow already exists, we draw a different destination). The specific routes corresponding to the truck flows are obtained by applying a shortest-path algorithm. The truck driver volume parameters f_q are calculated as $f_q = \eta_{O_q}\eta_{D_q}$, and normalized such that $\sum_{q\in Q} f_q = 100 n_{OD} n_r$ (hence, the average truck driver volume equals 100). Here,

η_{O_q} and η_{D_q} denote the population at the origin and destination of flow q, respectively, which are drawn from a uniform distribution on $[0, 1]$. Finally, we generate the patient volume parameters as $d_k = d_k^f + d_k^v$, where $d_k^f = 10$ represents the fixed patient volume and d_k^v the patient volume that depends on the flow volume passing this location, f_k. The first reflects the fact that North Star only selects potential locations at which the expected patient volume exceeds a given threshold. We draw d_k^v from a Gamma distribution with shape parameter f_k and scale parameter 1, and normalize them such that $\sum_{k \in K} d_k^v = 20|K|$ afterwards (hence, the average patient volume equals 10 + 20).

To generate the clustered instances, we start with a 6×6 grid on $[0, 1000]^2$. Among the 36 cells, we randomly select n cells to be assigned a total number of n_{OD} O-D nodes to. Here, the number n is drawn from a discrete uniform distribution on the set $\{3, 4, ..., 8\}$. The assignment of the O-D nodes is performed by distributing them over the n cells first, and by assigning them a random position within their assigned cell next (according to a continuous uniform distribution). The number of nodes assigned to one of the n cells is drawn from a discrete uniform distribution on the interval on $\{0, 1, ..., n_u\}$, where n_u represents the number of unassigned nodes. Next, we generate a road network. In each of the cells in the grid we generate 30 road network nodes, according to a continuous uniform distribution. Next, we generate an initial set of road network arcs by connecting each node (including the O-D nodes) to its closest 5 neighbors. If the graph is not yet connected afterwards, we add for each node an arc connecting this node with the closest neighbor it was not yet connected to. This process is repeated until the graph is connected. Next, we choose $|KP|$ of the road network nodes to become potential RWC locations, and transform the graph into a graph consisting of only RWC locations (O-D nodes and potential RWC locations) and arcs connecting them (i.e., an arc between two of these locations is introduced if there exists a path between them such that no other RWC location is passed). All other parameters are determined in the same way as done for the non-clustered instances.

Chapter 5

Forecasting Human African Trypanosomiasis Prevalences from Population Screening Data using Continuous Time Models[1]

5.1 Introduction

Human African trypanosomiasis (HAT), also known as sleeping sickness, is a parasitic disease that is caused by two sub-species of the protozoa Trypanosoma brucei: Trypanosoma brucei gambiense (gambiense HAT) and Trypanosoma brucei rhodesiense (rhodiense HAT). The infection causing the disease is transmitted from person

[1] Apart from several minor adaptations, this chapter is a direct copy of the article: H. de Vries, A.P.M. Wagelmans, E. Hasker, C. Lumbala, P. Lutumba, S.J. de Vlas, J.J. van de Klundert (2016). *Forecasting Human African Trypanosomiasis Prevalences from Population Screening Data using Continuous Time Models.* Published in PLoS Computational Biology.

to person through the tsetse fly. It is estimated that there were 20 000 cases in the year 2012 (WHO, 2015b) and that 70 million people from 36 Sub-Saharan countries are at risk of HAT infection (Simarro et al., 2012; WHO, 2015a).

Our work focuses on gambiense HAT, which represents 98% of all HAT cases WHO (2015a). Gambiense HAT, which we will refer to as "HAT" from now on, is a slowly progressing disease and is fatal if left untreated. In the first stage of the disease, symptoms are usually absent or non-specific (Brun et al., 2010). The median duration of this stage is about 1.5 years (Checchi et al., 2008). By the time patients arrive at a healthcare provider, the disease has often progressed to the neurological phase, which causes severe health problems. In addition, this treatment delay increases the rate of transmission, since an infected patient is a potential source of infection for the tsetse fly (Brun et al., 2010; Fevre et al., 2006). Therefore, active case finding and early treatment are key to the success of gambiense HAT control (WHO, 2013; Hasker et al., 2010).

The current case finding strategy uses mobile teams that travel from village to village to conduct exhaustive population screening (Mpanya et al., 2012; Brun et al., 2010; Hasker et al., 2010). For example, 35 mobile teams are active in the Democratic Republic of the Congo (DRC). Because this strategy has considerably reduced disease prevalence in several African countries (Rock et al., 2015; Fevre et al., 2006; Moore and Richer, 2001; Paquet et al., 1994), the disease is no longer perceived as a major threat. Consequently, donors are now scaling down their financial commitments (Hasker et al., 2010). This, however, poses a serious risk to the control of HAT. The disease tends to re-emerge when screening activities are scaled down, bringing about the risk of a serious outbreak, as shown by an epidemic in the 1990s (Brun et al., 2010; Moore and Richer, 2001; Van Nieuwenhove et al., 2001). For example, the number of cases in 1998 is estimated to have exceeded 300 000 (WHO, 2015a).

In order to minimize the risk of re-emergence when resources are scaled down, and in order to eliminate and eradicate the disease, maximizing the *effectiveness* of the control programs is crucial. Mpanya et al. (2012) suggest that the effectiveness of population screening is determined by (among others) the management and planning of the mobile teams. Planning decisions – which determine *which villages* to screen, and at what *time interval* to screen them – have a direct impact on the risk and

the magnitude of an outbreak. Existing literature does not address these issues, as highlighted by the WHO (2015b), and a wide variety of screening intervals have been applied in different control programs (Ruiz et al., 2002; Simarro et al., 1990; Paquet et al., 1994). To optimize the planning decisions, it is of key importance to be able to predict the evolution of the HAT prevalence level in the villages at risk. This allows decision makers to assess the relative effectiveness of a screening round in these villages and to prioritize the screening rounds to be performed.

However, practical tools for predicting HAT prevalence appear to be lacking. Existing models for HAT are mostly based on differential equations, describing the *rate of change* for the HAT prevalence level among humans and flies as a function of the prevalence levels among humans and flies (some models also include an animal reservoir) (Stone and Chitnis, 2015; Funk et al., 2013; Medlock et al., 2013; Artzrouni and Gouteux, 2001; Chalvet-Monfray et al., 1998; Jusot et al., 1995; Rogers, 1988). As the information needed to use such models – e.g., the number of tsetse flies in a village – is not available on the village level, using these models for prediction is impractical.

This chapter therefore sets out to develop practical models describing and predicting the expected evolution of the HAT prevalence level in a given village, based on historical information on HAT cases and screening rounds in that village. The main difference with the models mentioned in the previous paragraph is that our models make no assumptions about the causal factors underlying the observed prevalence levels: the "inflow" of newly infected persons and the "outflow" of infected persons by cure or death. Instead, we just consider data on the net effect of these two processes – the evolution of the prevalence level – and fit five different models to this. To analyze the predictive performance of these models, we make use of a dataset describing screening operations and HAT cases in the Kwamouth district in the DRC for the period 2004 - 2013.

Furthermore, we use one of the models to analyze the fixed frequency screening policy, which assigns to each village a fixed time interval for consecutive screening rounds. Specifically, we investigate screening frequency requirements for reaching *elimination* and *eradication*. Here, eradication is defined as "letting the expected

prevalence level go to zero in the long term", and elimination is defined as "reaching an expected prevalence level of one case per 10 000".

Our chapter thereby contributes to the branch of research on control strategies for HAT. Next, we list several other works that are highly related to ours. The effectiveness of active case finding operations is analyzed by Robays et al. (2004), who define "effectiveness" as the expected fraction of cases in a village which will eventually get cured as a result of a screening round in that village. The papers by Stone and Chitnis (2015), Chalvet-Monfray et al. (1998), and Artzrouni and Gouteux (1996) introduce differential equation models to gain structural insights on the effectiveness of combinations of active case finding and vector control efforts and on the requirements for eradicating HAT. The effect of active case finding activities is modeled through a continuous "flow" of infected individuals into the susceptible compartment. Since we explicitly model the timing and the effects of a screening round, this is one of the main differences with our paper. Finally, Rock et al. (2015) study the effectiveness of screening and treatment programs and the time to elimination using a multi-host simulation model. Their paper, however, considers the screening frequency as a given, whereas we consider the effects of changing this frequency. Furthermore, we propose models for predicting prevalence on a village level, whereas their model implicitly assumes all villages to be homogeneous.

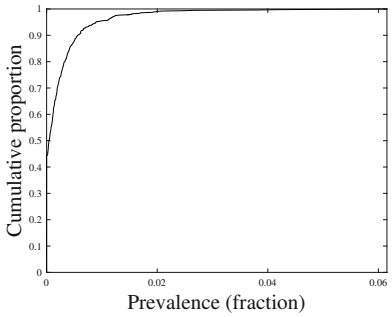
5.2 Materials and Methods

5.2.1 Data

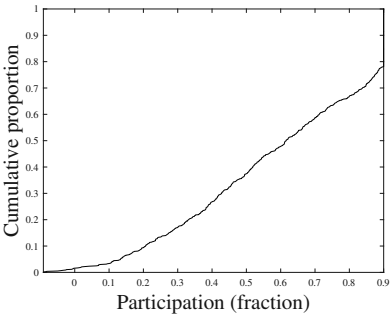
Our dataset consists of information on screening operations in the period 2004 - 2013 in the health zone Kwamouth in the province Bandundu. The raw data were cleaned up based on the rules described in Appendix 5.A. The number of villages in the dataset equals 2324, and 143 of these villages were included in the data analysis based on three criteria: (1) the number of screening rounds recorded was at least two, (2) at least one case has been detected over the time horizon, and (3) at least one record of the number of people screened during the operation was available. The first condition is necessary to enable modeling the prevalence level observed in a

given screening round as a function of past observed prevalence levels, and the third condition is necessary for estimating prevalence itself. We estimate the prevalence level in a village at the time of a screening round as the number of cases detected in that round over the number of people participating in that round. Furthermore, lacking population size data, we estimate the population of a village as the maximum number of people participating in a screening round reported for that village. Though our dataset also contains cases identified by the regular health system in between successive screening rounds, these do not yield (direct) estimates of prevalence levels in the corresponding villages, as required by the models proposed in the next section. We therefore focus on the active case finding data only.

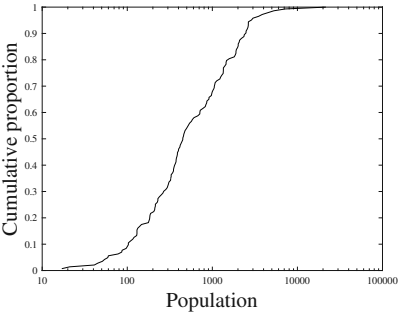
The total number of screening rounds reported for the 143 villages included equals 766 (on average 5.4 per village). Figure 5.1 shows cumulative distributions of the observed prevalence level in these screening rounds (mean 0.0055, median 0.0011, standard deviation 0.0121), the time interval between each pair of consecutive screening rounds (mean 1.28, median 1.00, standard deviation 1.03), the estimated population for each village (mean 1073, median 450, standard deviation 2046), and the participation level in the screening rounds (mean 0.69, median 0.72, standard deviation 0.27). Note that the relatively large number of observations with a participation level of 100% is due to the method used to estimate the population sizes.



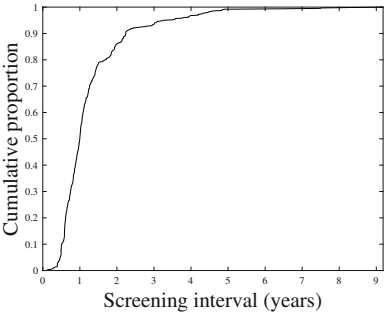
(a) Observed prevalence level in 766 screening rounds for the 143 villages in Kwamouth.



(b) Participation in the 766 screening rounds for the 143 villages in Kwamouth.



(c) Estimated number of people living in the 143 villages in Kwamouth.



(d) Time between 623 consecutive screening rounds for the 143 villages in Kwamouth.

Figure 5.1: Descriptive statistics of the dataset.

5.2.2 Prediction Methods

Before we propose our prediction methods, we introduce some notations. A table of the most important notations used in this chapter can be found in Appendix 5.G. Let $\mathbf{s}_v = \{s_{v1}, s_{v2}, ...\}$ denote the vector of screening time intervals for village v, where s_{v1} denotes the time between the start of the time horizon and the first screening for this village, s_{v2} denotes the time between the first and the second screening, and so on. The time at which the n^{th} screening is performed is given by $S_{vn} = \sum_{m \leq n} s_{vm}$ and the participation fraction in this screening round is denoted by p_{vn}. Parameter

N_v represents the population size of village v. Furthermore, let $\mathbf{i}_v$ represent historical information on HAT cases in this village: the numbers of cases detected during past screening rounds. We model the *expected* prevalence level at time t in village v as a function $f_v(\cdot)$ of $\mathbf{s}_v$, $\mathbf{i}_v$, and some parameters β: $f_v(t, \mathbf{s}_v, \mathbf{i}_v, \beta)$. Note that the expected prevalence level is a *latent*, i.e. unobserved, variable, and that the observed prevalence level, $x_v(t)$, generally deviates from the expected value. We measure prevalence levels $f_v(\cdot)$ and $x_v(\cdot)$ as *fractions* and represent the difference between the expected and observed prevalence level in village v by the random variable ε_v:

$$x_v(t) = f_v(t, \mathbf{s}_v, \mathbf{i}_v, \beta) + \varepsilon_v \tag{5.1}$$

Time series models such as discrete time ARMA, ARIMA or ARIMAX models seem to be the most popular methods for predicting prevalence (or incidence) (see e.g. Abeku et al., 2002; Allard, 1998). These models describe the prevalence level at time t as a linear function of the prevalence levels at time $t-1, t-2, ...$ and (optionally) some other variables. Their applicability in our context is however limited. Discrete time models require estimates of the prevalence level at each time unit (e.g., each month), whereas information to estimate the HAT prevalence level is available only at moments at which a screening round is performed. Namely, many HAT patients are not detected by the regular health system, particularly if they are in the first stage of the disease (Hasker et al., 2010).

The class of continuous time models is much more suitable for analyzing data observed at irregularly spaced times. These models assume that the variable of interest, $f_v(t, \mathbf{s}_v, \mathbf{i}_v, \beta)$, follows a continuous process, defining its value at each $t > 0$. The next subsections propose five continuous time models for predicting HAT prevalence levels. We again note that models describing the causal processes determining the observed prevalence levels in detail (e.g., by explicitly modelling disease incidence, passive case finding, death and cure) may be most intuitive, but require data that are not available on a village level. Therefore, to safeguard their relevance for practical application, the variables we include are only those that are available on a large scale. This does not imply that our models neglect the causal processes. In-

stead, they are to some extent accounted for in an implicit way by fitting the models to the observed prevalence levels.

Data that are typically available at village level are numbers of HAT cases found during screening rounds and the times of these screening rounds. For a given village, the first yields estimates of *past prevalence levels*, and the latter yield the *time intervals* between past screening rounds. We hypothesize that the current expected prevalence level at time t is related to past prevalence levels, past screening intervals, and in particular the time since the last screening round, which we denote by $\delta_v^-(t) = \min_n\{t - S_{vn} | S_{vn} \leq t\}$. Hence, we include (functions of) these variables in our models.

Linear regression models are very widely used in the world of forecasting (see e.g. Franses (1998)). Major advantages of these models are that they are easy to understand, to implement, to fit, and to analyze. Therefore, the first model we introduce is a linear model (model 1), which also serves as a benchmark for our more advanced models. This model describes the expected HAT prevalence in a given village as a function of the time since the last screening and past prevalence levels. Such linear model is, however, very vulnerable to a typical structure present in active case finding datasets. High past prevalence levels tend to increase the priority of screening a village, causing the time intervals between screening rounds to decrease. As a result, $\delta_v^-(t)$ is a highly "endogenous" variable. More formally, external variables (past prevalence levels) are correlated with both the dependent variable ($f_v(t)$) and the independent variable ($\delta_v^-(t)$), which makes it hard to quantify the (causal) relation between them. In response to this, we present four alternative models. Model 2 is a fixed effects model, which adds a dummy-variable for each village to the initial model. Model 3 is a (non-linear) exponential growth and decay model which is inspired by the SIS epidemic model. This model is being used extensively for modeling epidemics that are characterized by an initial phase in which the number of infected individuals grows exponentially, and a second phase in which this number levels off to a time-invariant *carrying capacity*. We refer to model 3 as the logistic model with a constant carrying capacity. Finally, model 4 is a less data dependent version of model 3 and model 5 is variant of model 3 in which the carrying capacity is allowed to vary over time.

Model 1: Linear Model. Let $x_v(S_{vn})$ and $\bar{x}_v(S_{vn})$ represent the observed prevalence level exactly at the n^{th} screening round in village v and the average observed prevalence level in the *three years prior to* the n^{th} screening round in village v, respectively. (For screening rounds in the first $T < 3$ years of our dataset, $\bar{x}_v(S_{vn})$ represents the average prevalence levels in these T years). The linear model (LM) describes the expected prevalence level at time t, $f_v(S_{vn} + \delta_v^-(t))$ as a function of the time since the last screening round in village v before time t, and past observed prevalence levels in that village:

$$\begin{aligned} x_v(S_{vn} + \delta_v^-(t)) &= f_v(S_{vn} + \delta_v^-(t)) + \varepsilon_v \\ &= \beta_1 \bar{x}_v(S_{vn}) + \beta_2 x_v(S_{vn}) + \beta_3 \sqrt{\delta_v^-(t)} + \beta_4 \delta_v^-(t) \\ &\quad + \beta_5 \bar{x}_v(S_{vn}) \sqrt{\delta_v^-(t)} + \beta_6 x_v(S_{vn}) \sqrt{\delta_v^-(t)} \\ &\quad + \beta_7 \bar{x}_v(S_{vn}) \delta_v^-(t) + \beta_8 x_v(S_{vn}) \delta_v^-(t) + \varepsilon_v \end{aligned} \tag{5.2}$$

We include the term $\sqrt{\delta_v^-(t)}$ into this model to incorporate the non-linear nature that characterizes epidemics: the (increase in the) number of cases typically levels off after an initial phase of fast growth. Furthermore, we include the cross terms based on the hypothesis that the prevalence level increases faster over time for villages which have higher past prevalence levels.

Model 2: Fixed Effects Model. Datasets about the HAT epidemic, such as the data set used in this study, show much similarity to *panel data* in the sense that both our data and panel data consist of observations at multiple points in time for multiple units (i.e., villages). The causality problem mentioned in the introduction of this section can be regarded as the problem that differences between units explain part of the regressors. A common way to deal with differences between units in panel data is to assume that there is a unit-specific, time-invariant *fixed effect* that contributes to the dependent variable. This is modeled by adding a constant for each unit to the regression model. Thereby, the model relates deviations from the fixed

effect to variations in the explanatory variables. In line with this, we consider the following regression model, which we denote by the fixed effects model (FEM):

$$\begin{aligned} x_v(S_{vn} + \delta_v^-(t)) &= f_v(S_{vn} + \delta_v^-(t)) + \varepsilon_v \\ &= \alpha_v + \beta_1 \bar{x}_v(S_{vn}) + \beta_2 x_v(S_{vn}) + \beta_3 \sqrt{\delta_v^-(t)} + \beta_4 \delta_v^-(t) \\ &\quad + \beta_5 \bar{x}_v(S_{vn}) \sqrt{\delta_v^-(t)} + \beta_6 x_v(S_{vn}) \sqrt{\delta_v^-(t)} + \beta_7 \bar{x}_v(S_{vn}) \delta_v^-(t) \\ &\quad + \beta_8 x_v(S_{vn}) \delta_v^-(t) + \varepsilon_v \end{aligned} \tag{5.3}$$

Here, α_v represents the fixed effect for village v. Note that the term α_v forms the only difference between the FEM and the linear model.

Model 3: Logistic Model with a Constant Carrying Capacity. Literature suggests that the development of the expected fraction of individuals infected with HAT grows exponentially in a first phase and that this fraction levels off to (or oscillates around and converges to) an equilibrium prevalence level in a second phase (Jusot et al., 1995; Rogers, 1988). We refer to the equilibrium prevalence level as the *carrying capacity*. A simple epidemic model that incorporates these effects is the SIS model, which describes the evolution of the expected prevalence level in village v by means of the following differential equation:

$$\frac{df_v(t)}{dt} = \kappa \cdot f_v(t) \left(1 - \frac{f_v(t)}{K_v}\right) \tag{5.4}$$

Here, as in the previous section, $f_v(t)$ denotes the expected fraction of people in population v who are infected at time t, K_v represents the carrying capacity of this village, and κ represents a parameter indicating the speed of convergence to the epidemic equilibrium. When the epidemic is in its initial phase, the growth rate of $f_v(t)$ is approximately $\kappa \cdot f_v(t)$, whereas this rate goes to 0 when $f_v(t)$ approaches the carrying capacity K_v. We propose to use the explicit solution to this differential equation to model the development of the expected HAT prevalence level in village v.

This solution is obtained by multiplying both sides of Equation (5.4) by dt, taking the integral on both sides, and regarding the previous screening (i.e., the one performed at time S_{vn}) as the beginning of the time horizon:

$$f_v(S_{vn} + \delta_v^-(t)) = \frac{K_v}{1 + A_{vn} \cdot e^{-\kappa \cdot \delta_v^-(t)}} \tag{5.5}$$

A_{vn} can be interpreted as a parameter determining the expected prevalence level *immediately after* the n^{th} screening round in village v: $A_{vn} = \frac{K_v}{f_v(S_{vn}^+)} - 1$, where S_{vn}^+ denotes this moment.

We fit the parameters in function (5.5) using the dataset described in the previous section. To keep the model simple, we assume that κ is constant over time and over villages (i.e., this parameter is an intrinsic property of the epidemic). Furthermore, we specify an interval of realistic values for κ based on the following reasoning. Note that κ represents the expected number of new infections per infected person per time unit in a population that is almost completely susceptible: $f_v(t) \to 0$. Now let r represent the yearly removal rate for HAT – the rate of progression to death or cure – and let R_0 denote the basic reproduction number for HAT – the average number of secondary cases induced by one case in an otherwise susceptible population. Since the expected disease duration of this one case is $\frac{1}{r}$ years, the average rate at which secondary infections are induced by this one case equals $\frac{R_0}{\frac{1}{r}} = r \cdot R_0$ per year. Under the assumption that this rate is constant over the disease duration, this implies that $\kappa = r \cdot R_0$. Since several analyses suggest that the value of R_0 is generally close to 1 in endemic regions (Rock et al., 2015; Funk et al., 2013; Davis et al., 2011), we hypothesize that $R_0 \in [1.0, 1.5]$. Furthermore, 95% confidence intervals for the stage 1 and stage 2 disease duration, as presented by Checchi et al. (2008), suggest that r lies in the interval $[0.22, 0.52]$. Based on these ranges we estimate that κ lies in the interval $[0.2, 0.8]$. We note that one could try to estimate village-specific or focus-specific ranges for R_0 (and hence for κ) using the next-generation matrix method, as proposed by Diekmann et al. (1990). Yet, as the corresponding data requirements are relatively large, we deem this to be practically unsuitable.

Furthermore, we hypothesize that K_v is highly related to past prevalence levels and past screening activities in village v. Higher prevalence levels indicate a higher carrying capacity. Furthermore, if two villages had equal prevalence levels, but if village 1 has been screened more frequently than village 2, this may indicate that the carrying capacity in village 1 is higher. In accordance with these hypotheses, we estimate K_v as:

$$K_v = \beta_1 + \beta_2 \tilde{\mu}_v \tilde{x}_v \tag{5.6}$$

Here, $\tilde{x}_v$ and $\tilde{\mu}_v$ denote the observed average prevalence level in village v during five consecutive years and the average screening frequency (screening rounds per year) in this village during these years, respectively. We take the last five years of data in our estimation sample (as we discuss in the next section, we split up the dataset in an estimation sample and a prediction sample). If only $T < 5$ years of data are available, we base $\tilde{x}_v$ and $\tilde{\mu}_v$ on these T years. Note that variables $\tilde{x}_v$ and $\tilde{\mu}_v$ are based on historical information and hence that these variables (and hence the estimated carrying capacity) are not affected by current screening frequency decisions. Furthermore, we note that these variables should be based on the same period of five years, since *combined information* on the screening efforts and resulting prevalence levels provides insight into the epidemic potential in a village.

The estimate of K_v provides one of the two inputs for determining A_{vn}. The second input is the expected prevalence level in village v at time S^+_{vn} (i.e., immediately after the n^{th} screening round). Under the assumption that infected individuals who are detected during a screening round are immediately "removed", the only people infected at time S^+_{vn} are those who did not participate in the screening round and those not detected by the diagnostic test. This suggests the following relation between the expected prevalence after and the expected prevalence before the n^{th} screening round in village v:

$$f_v(S^+_{vn}) = (1 - p_{vn} \cdot s)\, f_v(S^-_{vn}) \tag{5.7}$$

Here, p_{vn} and s denote the participation in the n^{th} screening in village v and the sensitivity of the diagnostic test, and S_{vn}^- denotes the moment right *before* the infected individuals were removed. We assume that $s = 0.925$ based on the review by Brun et al. (2010), and obtain the following estimator for A_{vn}:

$$A_{vn} = \frac{K_v}{f_v(S_{vn}^+)} - 1 = \frac{\beta_1 + \beta_2 \tilde{\mu}_v \tilde{x}_v}{(1 - p_{vn} \cdot s) f_v(S_{vn}^-)} - 1 \tag{5.8}$$

The model requires an assumption about the (unobserved) expected prevalence level at the beginning of the time horizon (i.e., at 01-01-2004). The only information we have about this level is that it is lower than the expected prevalence level at the first screening round. Furthermore, under the realistic assumption that the epidemic is in its "convex part", the expected prevalence level will have steeply increased between 01-01-2004 and the first screening round, suggesting that there is a substantial difference between the two. Therefore, taking $\tilde{x}_v$ as an estimate of the expected prevalence level at the first screening round and lacking further information, we choose to estimate $f_v(0)$ as half the average *observed* prevalence level during five years: $0.5\tilde{x}_v$. Section 5.3.2 discusses the sensitivity of our results with respect to this choice.

Substituting Equation (5.6) and Equation (5.8) into the definition of the prevalence level (5.5) we obtain the following model, which we refer to as the Logistic Model with a Constant Carrying Capacity (LMCCC), and which we fit by estimating the parameters κ, β_1, and β_2:

$$\begin{aligned} x_v(S_{vn} + \delta_v^-(t)) &= f_v(S_{vn} + \delta_v^-(t)) + \varepsilon_v \\ &= \frac{\beta_1 + \beta_2 \tilde{\mu}_v \tilde{x}_v}{1 + \left(\frac{\beta_1 + \beta_2 \tilde{\mu}_v \tilde{x}_v}{(1 - p_{vn} \cdot s) f_v(S_{vn}^-)} - 1 \right) \cdot e^{-\kappa \cdot \delta_v^-(t)}} + \varepsilon_v \end{aligned} \tag{5.9}$$

Note that the expected prevalence level before the n^{th} screening round, $f_v(S_{vn}^-)$, depends on the estimated values of κ, β_1, and β_2. In fact, by substituting its definition, Equation (5.9) can be written in the following form (see Appendix 5.B):

$$x_v(S_{vn} + \delta_v^-(t)) = \frac{\beta_1 + \beta_2 \tilde{\mu}_v \tilde{x}_v}{1 + \sum_{i \leq n} a_{vi} e^{-\kappa \cdot \delta_{vi}}} + \varepsilon_v \tag{5.10}$$

Here, a_{vi} and δ_{vi} denote some nonnegative constants.

Model 4: Restricted Logistic Model with a Constant Carrying Capacity. Estimates of historical prevalence levels $\tilde{x}_v$ are available on a very large scale. The national HAT control programs are a main source of these data, as they keep track of all HAT cases found. Alternatively, modeling studies have yielded estimates of the incidence levels in Sub-Saharan Africa at the level of detail of 1 km^2 (see Simarro et al. (2010, 2012)), which can be transformed into corresponding estimates of prevalence levels. Data to measure $\tilde{\mu}_v$, the historical screening frequency in village v, are however scarcer in general, as gathering these data brings about a significant administrative burden. It is therefore relevant to investigate the predictive performance of a model that only uses $\tilde{x}_v$ to estimate K_v instead of $\tilde{\mu}_v \tilde{x}_v$. We do so by fitting the following model, which we refer to as the restricted Logistic Model with a Constant Carrying Capacity (rLMCCC):

$$x_v(S_{vn} + \delta_v^-(t)) = \frac{\beta_1 + \beta_2 \tilde{x}_v}{1 + \left(\frac{\beta_1 + \beta_2 \tilde{x}_v}{(1 - p_{vn} \cdot s) f_v(S_{vn}^-)} - 1 \right) \cdot e^{-\kappa \cdot \delta_v^-(t)}} + \varepsilon_v \tag{5.11}$$

Model 5: Logistic Model with a Varying Carrying Capacity. The models presented in the previous subsections implicitly assume that the carrying capacity (i.e., the upper bound on the expected prevalence level) remains the same over time. However, due to possible changes in conditions that affect the disease dynamics – such as vegetation, the number of flies and humans in and around a village, passive case finding activities – it is realistic to assume that the carrying capacity may vary over time. Furthermore, the epidemic potential may also vary due to variations in susceptibility to or tolerance of HAT infection, as argued by Welburn et al. (2016). Based on this assumption, we propose a Logistic Model with a Varying Carrying Capacity (LMVCC):

$$x_v(S_{vn} + \delta_v^-(t)) = \frac{\beta_1 + \beta_2 \bar{\mu}_v(S_{vn})\bar{x}_v(S_{vn})}{1 + \left(\frac{\beta_1 + \beta_2 \bar{\mu}_v(S_{vn})\bar{x}_v(S_{vn})}{(1 - p_{vn} \cdot s) f_v(S_{vn}^-)} - 1\right) \cdot e^{-\kappa \cdot \delta_v^-(t)}} + \varepsilon_v \qquad (5.12)$$

The only change with respect to the logistic model with a constant carrying capacity is that $\tilde{x}_v$ and $\tilde{\mu}_v$ are replaced by $\bar{x}_v(S_{vn})$ and $\bar{\mu}_v(S_{vn})$, the average observed prevalence level in and the screening frequency for village v in the *three years prior to* the n^{th} screening round in that village, respectively.

5.2.3 Model Fitting

As HAT prevalence levels are very low, the variance of these levels is high, which enhances the chance that there are significant outliers among the observations. For example, no cases were detected in three out of four screening rounds performed in a village of 122 people, whereas five cases were detected in the 4^{th} round. This implies two things. First, observed prevalence levels will generally deviate significantly from expected prevalence levels. Second, we need to choose a technique for estimating the model coefficients that is robust with respect to outliers. Instead of Least Squares (LS) regression, one of the most commonly applied model fitting methods, we therefore use Least Absolute Deviations (LAD) regression to fit the model parameters, which is known to be relatively insensitive to outlying observations (Heij et al., 2004). An alternative technique would be to use a maximum likelihood estimation (MLE) approach based on a heavy-tailed probability distribution for the observed prevalence levels. In Appendix 5.D, we show the results obtained when assuming a Poisson, Beta-Binomial, or Negative Binomial distribution. Each of the MLE approaches is, however, clearly outperformed by the LAD regression approach.

The variance of the observed prevalence level strongly depends on the sample size. For example, under the assumption of an independent infection probability for each person, the variance is inversely proportional to the sample size. We therefore weight the fitting deviation $e_{vn} = f_v(S_{vn}) - x_v(S_{vn})$ for observation n for village v by weight $w_{vn} = \sqrt{N_v \cdot p_{vn}}$, yielding the following weighted LAD regression problem:

$$\min_{\beta} S_{abs}(\beta) = \sum_{(v,n)} w_{vn}|e_{vn}| \tag{5.13}$$

To deal with the risk of overfitting, we select the variables to be included in the models by means of a backward elimination method. This method initially includes all variables in the model and iteratively removes the least significant variable (if its p-value > 0.10) and estimates the model with the remaining variables. The algorithm stops as soon as all remaining variables are significant or if only one variable is left. We enforce that α_v and κ cannot be removed by the backward elimination method so as to preserve essential elements of the corresponding models. Hence, only parameters β_1-β_8 in models 1 and 2, and parameters β_1-β_2 in models 3-5 could be removed.

Finally, to test the predictive performance of the models, we split the data in an estimation sample (which we use for fitting the model) and a prediction sample. Specifically, for each of the 143 villages, we include the last screening round in the prediction sample, and include the others in the estimation sample. Next, we measure performance based on the mean of the prediction errors $ME = \frac{\sum_v e_{\widehat{vn}}}{|V|}$, indicating whether the predictions obtained by the model are *biased*, and based on two indicators for the amount of *explained variation* in the prevalence levels: the mean absolute error, $MAE = \frac{\sum_v |e_{\widehat{vn}}|}{|V|}$ and the mean relative error, $MRE = \sum_v \frac{|e_{\widehat{vn}}|}{\sum_v x_v(S_{\widehat{vn}})}$. Here, the index combination $\widehat{vn}$ indicates the last screening round for village v. The intuition behind the measures of explained variation is that they equal 0 if the predicted prevalence levels are exactly equal to the observed prevalence levels (i.e., the model perfectly explains the variation in the observed prevalence levels) and that their value increases when the absolute difference between predicted and observed levels increases (i.e. when the model explains less variation in observed prevalence levels). We use Matlab R2015b for the implementation of our methods.

5.3 Results

5.3.1 Fitted Models

Table 5.1 presents the coefficient estimates for the variables of the five presented models. The results for models 1 and 2 are very similar. Seven of the eight variables are identified as being non-significant by the backward elimination algorithm: The interaction terms, the long term prevalence level, the time since the last screening round, and the square root of the time since the last screening round. The resulting model provides a clear prediction method: The expected prevalence equals 24.5% of the prevalence level observed at the previous screening round (note, if this level was 0.0%, the estimated expected prevalence remains 0.0%) according to model 1, and equals 14.7% of this prevalence level plus a constant fraction α_v according to model 2. Hence, this model predicts that, in the absence of screening activities, the expected prevalence remains the same over time.

Variable	Coefficient name	Coefficient value	Std error	P-value
Model 1: Linear Model				
$x_v(S_{vn})$	β_2	0.245	0.035	0.000
Model 2: Fixed Effects Model*				
$x_v(S_{vn})$	β_2	0.147	0.055	0.004
Model 3: Logistic Model with a Constant Carrying Capacity				
$\tilde{\mu}_v\tilde{x}_v$	β_2	1.763	0.409	0.000
$\delta_v^-(t)$	κ	0.800	0.086	0.000
Model 4: Restricted Logistic Model with a Constant Carrying Capacity				
$\tilde{x}_v$	β_2	1.234	2.210	0.288
$\delta_v^-(t)$	κ	0.800	0.130	0.000
Model 5: Logistic Model with a Varying Carrying Capacity				
Constant	β_1	0.004	0.001	0.002
$\bar{\mu}_v(S_{vn})\bar{x}_v(S_{vn})$	β_2	0.330	0.157	0.018
$\delta_v^-(t)$	κ	0.800	0.123	0.000

* For sake of conciseness, the estimates of α_v are not given here.

Table 5.1: Estimated weighted LAD regression models.

The fitted models 3, 4, and 5 reveal a clear and intuitive relationship between screening frequency, prevalence, and carrying capacity: a larger historical prevalence indicates a higher carrying capacity, and facing an equal historical prevalence for a higher historical screening frequency indicates a higher carrying capacity. The constant term has been identified as non-significant for models 3 and 4 and as significant for model 5. To illustrate the typical output of models 3 and 4, Figure 5.2 shows the development of the expected prevalence levels for two villages over time (the lines), as well as the observed prevalence levels (stars and circles). Furthermore, Figure 5.3 depicts the carrying capacities for the 143 villages in Kwamouth, as estimated by the LMCCC model. Though data to validate these estimates are lacking, we note that they are in the same order of magnitude as prevalence levels found during screening rounds. The latter are usually between 1% and 5% in high or very high transmission areas, and exceed 10% in some extreme cases (Checchi et al., 2012; Pépin and Méda, 2001).

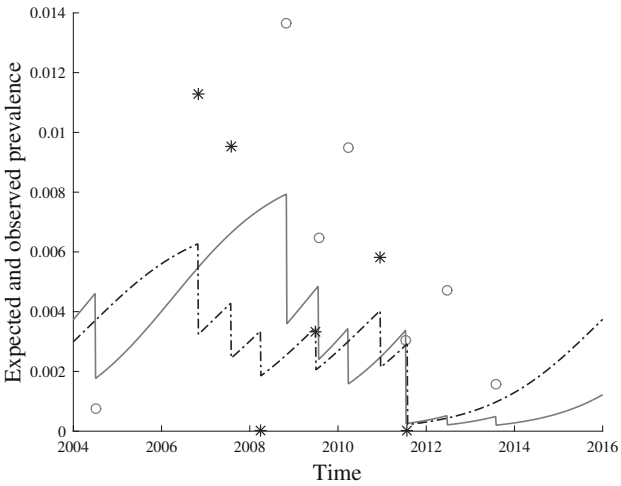


Figure 5.2: Evolution of the expected prevalence level over time for two villages according to the LMCCC model (lines) versus the observed prevalence levels (circles and stars). Village 1: solid line vs. circles. Village 2: dashed line vs. stars.

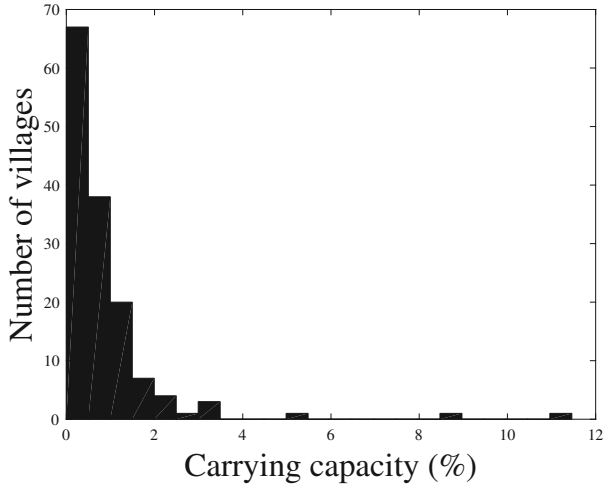


Figure 5.3: Histogram of estimated carrying capacities for the 143 villages in Kwamouth, as estimated by the LMCCC model. Measured as a percentage of the population.

5.3.2 Comparison of Models

As mentioned in Section 5.2.3, we measure the predictive performance of the different models in terms of the prediction bias and in terms of the amount of variation explained. Table 5.2 contains the values of the different indicators for each of the models and Appendix 5.H compares the prediction errors produced by the different models. These prompt several interesting observations.

	ME	MAE	MRE
Model 1: LM	−0.00186	0.00401	1.26
Model 2: FEM	−0.00105	0.00507	1.60
Model 3: LMCCC	−0.00094	0.00476	1.50
Model 4: rLMCCC	**0.00047**	0.00600	1.89
Model 5: LMVCC	−0.00176	**0.00363**	**1.14**

Table 5.2: Comparison of the predictive performance of the five models in terms of mean errors (ME), mean absolute errors (MAE), and mean relative errors (MRE). The best indicator values are in bold.

First, the prediction bias ranges from 0.47/1 000 (rLMCCC model) to -1.86/1 000 (LM model). Given that the average observed prevalence in the 766 screening rounds in our dataset equals 5.5/1 000, we consider the biases of the LM model and the LMVCC model as quite substantial. Yet, this may be very well explained by the highly variable character of the HAT epidemic. A small number of outbreaks may substantially shift the average observed prevalence level. For example, without the four most negative prediction errors, the prediction bias for the LM model would be only -0.69/1 000.

Second, the LM model performs relatively well in terms of explained variation. Yet, we see two vulnerabilities of this model: (1) as discussed before, this model is likely to be hampered by endogeneity, inducing a potential bias in the coefficient estimates, and (2) the variation in the screening intervals is relatively small for the villages with the highest endemicity levels in our sample, as these villages are screened almost every year. When there is little variation in $\delta_v^-(t)$, the true effects of variations might not become visible. These two fundamental vulnerabilities may very well explain why (a function of) $\delta_v^-(t)$ has not been identified as significant for the LM model. As a result, this model unrealistically predicts that the value of the expected prevalence level remains the same over time in the absence of screening activities, contrasting with vast historical evidence.

The same vulnerabilities apply to the FEM model, which also provides a counter-intuitive relation between the expected prevalence and $\delta_v^-(t)$. On top of that, its predictive power is relatively low, which could be explained by the fact that, for many villages, there is insufficient data to estimate the fixed effect accurately.

As variants of the logistic model already fix the structure of the relationship between $\delta_v^-(t)$ and $f_v(S_{vn} + \delta_v^-(t))$ based on epidemiological insights, these models do not suffer from the vulnerabilities mentioned above. We therefore consider these models to have most potential for accurately predicting HAT prevalence levels in general (i.e., in any region and for any time horizon). Among the three logistic model variants, model 3 (LMCCC) performs reasonably well in terms of both criteria. Model 5 (LMVCC) has a substantial prediction bias, but performs best in terms of explained variation. Though model 4 (rLMCCC) performs best in terms of prediction bias, it performs very weakly in terms of explained variation.

Hence, among the logistic model variants, there is no clear winner when both criteria are assigned equal importance. For planning decisions, however, we consider model 5 to be most suitable, followed by model 3. The reason is that, in contrast with prediction bias, explained variation indicates the ability to identify differences in expected prevalence levels between villages, as required for effective planning decisions. Hence, identifying an effective *prioritization* of the different villages will be more important than obtaining unbiased estimates of the resulting *prevalence* levels.

The sensitivity level s is known to differ between regions (Truc et al., 2002). Furthermore, the population size of a village had to be estimated, which induces a potential bias in the participation level estimates. These issues beg to question the robustness of our results on the logistic model variants (note, models 1 and 2 are not affected by this as these do not use these parameters). Appendix 5.E shows the results of a sensitivity analysis, which largely confirm our findings. In all scenario's analyzed, model 5 remains best in terms of explained variation, followed by model 3, and models 3 and 4 outperform model 5 in terms of prediction bias. Another assumption that questions the robustness of our results is the one about the expected prevalence level at the beginning of the time horizon (i.e., at 01-01-2004). Appendix 5.F provides the results of a sensitivity analysis on this assumption. Again our main findings remain the same.

5.3.3 Analysis of the Fixed Frequency Screening Policy

In the previous section we argue that, among the models analyzed in this chapter, variants of the logistic model have most potential for accurately predicting HAT prevalence levels in general. In this section we demonstrate the applicability of one of these model variants to analyze the effectiveness of screening operations. In particular, since information on the development of the carrying capacities is lacking, as required by the LMVCC model, and since we consider the predictive performance of model 3 superior to that of model 4, we choose to use the LMCCC model as a basis for this analysis. We do note that the *theoretical results* presented here also hold for model 4 and, if the carrying capacity remains constant, for model 5 also.

Our analysis will concentrate on the fixed frequency screening policy. This policy assigns to each village a fixed time interval for consecutive screening rounds based on the village's characteristics. As the policy is relatively easy to understand and implement, it has been the basis for guideline documents for HAT control. For example, the WHO recommends a screening interval of one year for villages reporting at least one case in the past three years, and an interval of 3 years for villages that did not report a case in the last three years, but did report at least one case during the past five years (WHO, 2015b).

In the first part of this section, we mathematically analyze the impact of a fixed screening policy for a given village and investigate the screening frequency required to *eradicate* HAT in that village. As mentioned in the introduction, we define that HAT is eradicated in the long term if the expected prevalence level goes to zero in the long term.

A shorter term objective is to *eliminate* HAT, where elimination is defined as having at most one new case per 10 000 persons per year (WHO, 2013, 2015b). For example, the WHO's roadmap towards elimination of HAT states the aim to eliminate (gambiense) HAT as a public health problem by 2020 – which is defined as having less than one new case per 10 000 inhabitants in at least 90% of the disease foci (WHO, 2015b) – and to reach worldwide elimination by 2030. The second part of this section presents analytical results about the time needed to reach elimination and about the screening frequency requirements for reaching elimination *within a given time frame*. As our models consider expected prevalence instead of incidence, we redefine elimination as "reaching an expected prevalence level of one case per 10 000". We argue that the times and efforts required to reach this elimination target are practically suitable lower bounds on the times and efforts needed to reach the WHO's targets. First, incidence and prevalence levels are argued to be "comparable" for HAT if mobile units visit afflicted areas infrequently (Stone and Chitnis, 2015). If mobile teams visit the areas more frequently, incidence will only become larger compared to prevalence and the prevalence level target will be easier to achieve than the incidence level target (e.g., under the assumption that the fraction of flies infected is proportional to the fraction of humans infected, this follows directly from the epidemic model presented by Rogers (1988)). Second, even if the *expected* prevalence

level is below the defined threshold level, the intrinsic variability of the HAT epidemic may induce an *actual* prevalence level that exceeds this threshold.

Throughout this section, we consider an imaginary village with a constant carrying capacity K. (For sake of conciseness we omit the subscript v in this section). Furthermore, we assume a constant participation level $p_{vn} = p$, $0 < p < 1$, and a fixed screening interval τ. The expected prevalence level at the beginning of the time horizon is denoted by $f(0)$, $f(0) > 0$. Finally, recall that s, $0 < s < 1$, denotes the sensitivity level.

Screening Requirements for Eradication. Let $\bar{f}_{n,n+1}$ denote the expected average prevalence level faced between screening rounds n and $n+1$. In Appendix 5.B, we prove the following results, which imply that the expected average HAT prevalence level will go to zero if the screening interval τ is at most $\frac{-\log(1-p\cdot s)}{\kappa}$ years, and that it will go to some strictly positive level otherwise:

Proposition 5.1. *If* $\tau \leq \frac{-\log(1-p\cdot s)}{\kappa}$, *then* $\lim_{n\to\infty} \bar{f}_{n,n+1} = 0$.

Proposition 5.2. *If* $\tau > \frac{-\log(1-p\cdot s)}{\kappa}$, *then* $\lim_{n\to\infty} \bar{f}_{n,n+1} = K\left(\frac{\log(1-p\cdot s)}{\kappa\cdot\tau} + 1\right)$.

A first insight provided by these propositions is that the eradication threshold for the screening interval strongly depends on p and s; The level of participation in the screening activities and the sensitivity of the diagnostic test. To illustrate this, Figure 5.4 shows the screening interval required for eradication for a range of values of the participation level and the sensitivity level. For example, if both values equal 70%, the required screening interval is 0.84 years (approximately once per 10 months). In case of participation and sensitivity levels of 99%, the required interval equals 4.9 years. This reveals the tremendous benefits of maximizing these values.

A second main insight is that annual screening, as currently recommended for endemic areas (WHO, 2015b), only leads to eradication if the case detection fraction – defined as participation times sensitivity – exceeds 55%. For example, this is realized when the participation and sensitivity levels are both at least 75%. Furthermore, under realistic current sensitivity and participation levels – 92.5% based on Brun et al. (2010), and 70% based on our dataset, respectively – eradication requires a screening frequency of at least once per 1.30 years (approximately 15 months).

A final insight is that the eradication threshold does not depend on the village's carrying capacity, meaning that the same upper bound on the screening frequency holds for all endemic areas. This can be understood by observing that changing the carrying capacity just changes the prevalence level with a constant factor at each point of time (see Equation (5.4)). Consequently, also the long term average prevalence changes with this constant factor, and hence remains zero if and only if the eradication threshold is met. We note that this does not imply that the disease burden, as measured by the number of person-years of illness, does not depend on the village's carrying capacity. In contrast, our results in Appendix 5.B reveal that, for a given screening policy, the disease burden scales linearly with the carrying capacity.

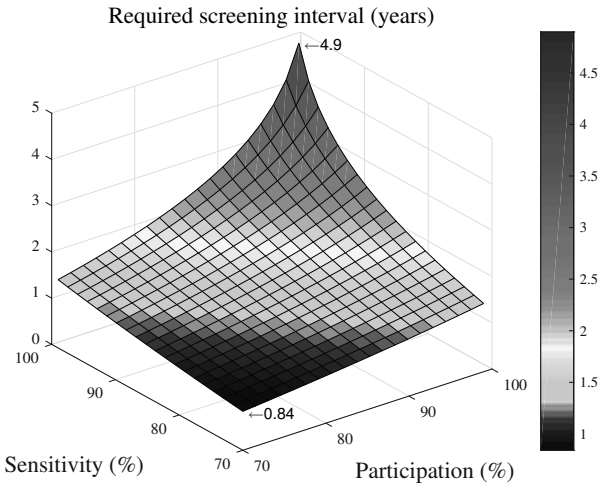


Figure 5.4: Screening interval required to reach elimination, for different levels of participation in the screening activities, and different levels of the sensitivity of the diagnostic test(s) performed.

Screening Requirements for Elimination. Let $\alpha = \frac{1}{1-p\cdot s}e^{-\kappa\cdot\tau}$, $\beta = \frac{p\cdot s}{1-p\cdot s}$, and $A_0 = \frac{K}{f(0)} - 1$. In Appendix 5.C, we prove the following result on the time till the expected prevalence level hits a predefined level C in case of a screening interval of length τ:

Proposition 5.3. *If $f(0) > C$ and $\tau \leq \frac{-\log(1-p\cdot s)}{\kappa}$, the expected prevalence level is smaller than or equal to C for the first time after screening round n^*, where*

$$n^* = \begin{cases} \left\lceil \alpha \log\left(\frac{\frac{K}{C} - 1 + \frac{\beta}{\alpha - 1}}{A_0 + \frac{\beta}{\alpha - 1}} \right) \right\rceil & \text{if } \alpha > 1 \\ \left\lceil \frac{\frac{K}{C} - 1 - A_0}{\beta} \right\rceil & \text{if } \alpha = 1 \end{cases} \tag{5.14}$$

This result enables us to estimate the time needed to reach the expected prevalence level of 1/10 000 for a given village. For example, let us consider a "typical village" with a current long term prevalence level ($\tilde{x}$) of 5/1 000 and let us assume that $\tilde{\mu} = 1$ (the median historic screening frequency in our dataset) and that $f(0) = 0.5\tilde{x}$. Next, let us use the fitted LMCCC model described in Table 5.1 to estimate the village's carrying capacity. Then, if annual screening is maintained and under the estimated current sensitivity and participation levels, the estimated time to elimination is 10 years.

Figure 5.5 provides some more insight into the relationship between the case detection fraction, the screening interval and the number of years needed to reach elimination. We see that the time to elimination ranges from 0.5 years (detection fraction: 99%, screening interval: 6 months) to 15 years (detection fraction: 75%, screening interval: 1.5 years) and that the detection fraction has a substantial impact on the time to elimination. For example, in case of a screening interval of 1.5 years, the time to elimination decreases from 15 years to 4.5 years when the case detection fraction increases from 75% to 90%. As explained, we consider the times shown here to be lower bounds on the times needed to reach the WHO's targets, which is in line with earlier findings (Stone and Chitnis, 2015; Rock et al., 2015).

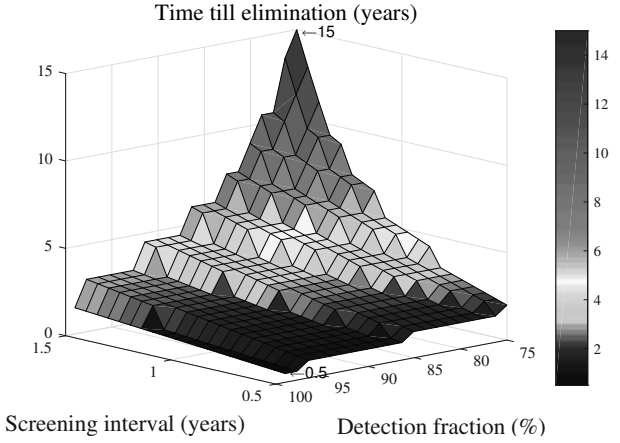


Figure 5.5: Time needed to reach elimination in a "typical village" (current prevalence level of 5/1 000), for different levels of the case detection fraction ($p \cdot s$) and different screening intervals.

Now, suppose that we would like to determine the screening interval needed to let the prevalence level cross the boundary C *within* T *years*. From the results stated in Appendix 5.B it follows that this is equivalent to finding n^*: the smallest number of screening rounds such that $A_{n^*} \geq \frac{K}{C} - 1$. Note that, under the fixed frequency policy, performing n screening rounds in T years implies a screening interval of $\frac{T}{n}$ years. In the following result, we assume that the first screening round is performed at time $\frac{T}{n}$ (and hence that the last is performed at time T):

Lemma 5.1. *Given that* $\tau = \frac{T}{n} \leq \frac{-\log(1-p\cdot s)}{\kappa}$, *then* $\lim_{n\to\infty} A_n = \infty$ *and the sequence* $\{A_n\}$ *is monotonically increasing in* n.

This lemma justifies the use of a simple search algorithm for finding the number of screening rounds to perform in the next T years, given the target prevalence level C. For example, this number can be found by increasing the value of n till A_n exceeds the threshold value $\frac{K}{C} - 1$.

We use this approach to determine the screening interval required to reach elimination *within 5 years from now* for different levels of the case detection fraction and different current long term prevalence levels ($\tilde{x}$) of an imaginary village. Again,

we use the fitted LMCCC model described in Table 5.1 for estimating the village's carrying capacity and assume that $\tilde{\mu} = 1$ and that $f(0) = 0.5\tilde{x}$.

Figure 5.6 depicts the results. We see that the number of screening rounds needed to reach elimination within five years ranges from 1 (i.e. $\tau = 5$) to 11 (i.e. $\tau = 0.45$). Another observation is that *annual* screening only leads to elimination within five years if the case detection fraction is very high – roughly above 75% – or if the current prevalence level is very low – roughly below 1/1 000. For the reason stated before, we consider this to be a lower bound on the efforts needed to reach the WHO's targets within five years.

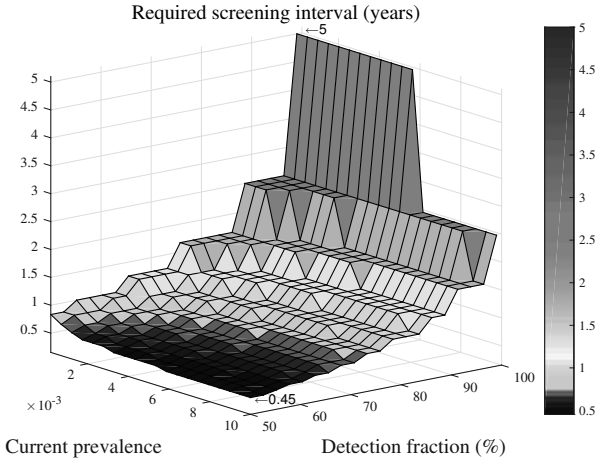


Figure 5.6: Screening interval required to reach elimination within five years, for different levels of the case detection fraction ($p \cdot s$) and different current long term prevalence levels ($\tilde{x}$).

5.4 Discussion and Conclusions

This chapter introduces and analyzes five models for predicting HAT prevalence in a given village based on past observed prevalence levels and past screening activities in that village. Based on the quality of prevalence level predictions in 143 villages in Kwamouth (DRC), and based on the theoretical foundation underlying the models,

we conclude that variants of the logistic model – a model inspired by the SIS model – are most practically suitable for predicting HAT prevalence levels. Sensitivity analyses show that this conclusion is very robust with respect to assumptions about participation levels, the sensitivity of the diagnostic test, or the initialization value of the prevalence curves are violated. Second, we demonstrate the applicability of one variant of the logistic model to analyze the effectiveness of the fixed frequency screening policy, which assigns to each village a fixed time interval for consecutive screening rounds.

Due to the intrinsic variability of the HAT epidemic, observed prevalence levels will generally deviate significantly from predicted prevalence levels. We strongly believe, however, that this does not render predictions worthless in the context of planning decisions. In contrast, a major contribution of our models is that they indicate the *expected* disease burden in different villages and can hence be applied to develop planning policies that aim to minimize the total *expected* disease burden for the villages considered.

Our analysis of the fixed frequency screening policy reveals that *eradication* of HAT is to be expected in the long term when the screening interval is smaller than a given threshold. This threshold strongly depends on the case detection fraction: the fraction of cases who participate in the screening rounds *and* are detected by the diagnostic tests. Under current conditions, we estimate the threshold to be approximately 15 months. This suggests that annual screening, as recommended by the WHO for endemic areas, will eventually lead to eradication. More specifically, our model predicts that annual screening will lead to eradication if the case detection fraction exceeds 55%.

The logistic model also reveals expressions for the time needed to reach the more short term target of *eliminating* HAT and for the screening interval required to eliminate HAT *within a given time frame*. These suggest that it takes 10 years to eliminate HAT in a village or focus with a prevalence of 5/1 000 (under current conditions and annual screening). Furthermore, we estimate that it is only feasible to reach elimination within five years if the case detection fraction is very high – roughly above 75% – or if the current prevalence level is very low – roughly below

1/1 000. We argue that these figures are practically suitable lower bounds on the time or efforts needed to reach the WHO's targets for elimination.

Our results on requirements for eradication or elimination are based on a deterministic model, which begs to question their validity for reality, where events are stochastic. We note, however, that we model the *expected* behavior of a stochastic system, and hence that our results also hold *in expectation* for the stochastic system. On the other hand, we acknowledge that our models are not perfect. For example, we neglect interaction effects between neighboring villages. It would therefore be interesting and relevant to investigate whether our results can be reproduced by a validated simulation model.

A necessary condition for the applicability of our prediction models is that data about possible past HAT cases are available. Obviously, they thereby fail to identify endemic villages that have never been visited by a screening team. Different types of models, making use of different types of data – such as vegetation data (see e.g. Rogers (2000)) or data about cases in nearby areas – are needed for these villages. A second limitation of the models is that they have been tested on a set of villages for which the screening interval was relatively short. Consequently, it is hard to establish the behavior of the epidemic if much larger screening intervals are used or if screening operations would be abandoned. Since datasets that do reveal this behavior are lacking, fitting the models to an appropriate dataset obtained by a validated simulation model would be a promising direction for future research.

We consider it likely that the speed of convergence to an epidemic equilibrium, as indicated by the parameter κ, differs per disease focus. This may be due to differences in epidemiological conditions, such as the presence of an animal reservoir and the specific tsetse subspecies living in a focus. These differences induce a need for developing and analyzing a cost-effective control strategy that takes differences between foci into account, as also recognized by Simarro et al. (2013). Repeating our analyses for different disease foci to investigate differences in requirements for elimination and eradication would therefore be highly relevant.

Our analyses revealed that the effectiveness of screening operations strongly depends on the participation level in the screening activities and the sensitivity of the diagnostic tests. Using screening algorithms with a higher sensitivity or case find-

ing procedures that increase participation levels therefore seems to be very effective. For example, Robays et al. (2004) list several combinations of diagnostic tests with different sensitivity levels, and a more acceptable case finding procedure is currently being piloted in the DRC. Using our model to investigate the cost-effectiveness of different control strategies, defining the screening frequencies as well as the screening procedures and diagnostic tests used, would be a promising direction for future research.

The results on the fixed frequency screening policy are all concerned with the screening frequency or the time needed to reach certain *targets*. While at a strategic level, these results might be interesting, at a tactical or operational level, screening frequency decisions are constrained by resource availability. Consequently, policies providing effective, acceptable, and practically suitable recommendations on how to allocate *available screening capacity* over the villages at risk are of much higher relevance at these levels. Literature about this allocation problem is however absent, as also observed by the WHO (2015b), so that future research addressing this topic is much needed.

Appendices

5.A Notes Data Cleanup

The following rules were used for cleaning up the database:

- If only the year of screening was reported, assume that it took place at July 1^{st}.
- If the month of screening was not reported, assume that it took place in July.
- If the day of screening was not reported or invalid (e.g., 0), assume that it took place at the 1^{st} day of the month.
- Only select data between 01-01-2004 and 31-12-2013 (there are some data about 2014, but these are incomplete).

- If the dates of two screening rounds are at most three months apart, we regard them as one screening round.
- If two screening rounds are entered in the same cell ($>$ 3 months apart), but the number of people screened is entered only once, assume that the number of people screened is the same for both rounds.
- If a screening round took multiple days, choose the first day as the screening date.
- If the number of people screened in a screening round is not reported, assume it to be equal to the average number of people screened in the rounds for which this number *is* reported.

5.B Proofs on Screening Requirements for Eradication

For sake of conciseness we omit the subscript v in this section. Before we state our proofs, let us make some observations about the expected prevalence level immediately after the n^{th} screening round:

$$f(S_n^+) = \frac{K}{1 + A_n} \tag{5.15}$$

Let us specifically focus on the parameter A_n. Note that a larger value of A_n implies a lower prevalence level after the screening round and vice versa. Recall that the impact of a screening round is estimated as (see main text):

$$f_v(S_{vn}^+) = (1 - p_{vn} \cdot s)\, f_v(S_{vn}^-).$$

This implies for any $n \geq 1$:

$$A_n = \frac{K}{f(S_n^+)} - 1 \tag{5.16}$$

$$= \frac{K}{(1 - p \cdot s) \cdot f(S_n^-)} - 1 \tag{5.17}$$

$$= \frac{K}{(1 - p \cdot s) \cdot \frac{K}{1 + A_{n-1} \cdot e^{-\kappa \cdot \tau}}} - 1 \tag{5.18}$$

$$= \alpha \cdot A_{n-1} + \beta \tag{5.19}$$

$$= \alpha^n \cdot A_0 + \sum_{i=1}^{n} \alpha^{n-i} \cdot \beta \tag{5.20}$$

$$= \begin{cases} \alpha^n \cdot A_0 + \beta \frac{\alpha^n - 1}{\alpha - 1} & \text{if } \alpha \neq 1 \\ A_0 + n \cdot \beta & \text{if } \alpha = 1 \end{cases} \tag{5.21}$$

Here, $\alpha = \frac{1}{1-p \cdot s} e^{-\kappa \cdot \tau}$, $\beta = \frac{p \cdot s}{1-p \cdot s}$, and $A_0 = \frac{K}{f(0)} - 1$. Note that α, β, and A_0 are strictly positive. These observations and definitions allow us to prove the following results.

Lemma 5.2. *If* $\tau \leq \frac{-\log(1-p \cdot s)}{\kappa}$, *then* $\lim_{n \to \infty} A_n = \infty$ *and* A_n *is discretely convex and increasing in* n.

Proof. First, rewriting yields that the condition $\tau \leq \frac{-\log(1-p \cdot s)}{\kappa}$ is equivalent to the condition $\alpha = \frac{1}{1-p \cdot s} e^{-\kappa \cdot \tau} \geq 1$. The result now follows immediately from Equation (5.21) and by the observations that $A_0 > 0$ and $\beta > 0$. If $\alpha > 1$ these imply that:

$$\lim_{n \to \infty} A_n > \lim_{n \to \infty} \alpha^n A_0 = \infty \tag{5.22}$$

Furthermore, if $\alpha = 1$, it holds that:

$$\lim_{n \to \infty} A_n = \lim_{n \to \infty} A_0 + n \cdot \beta = \infty \tag{5.23}$$

The discrete convexity of A_n is trivial for the case that $\alpha = 1$, since A_n is linear in n in this case. If $\alpha > 1$, the convexity follows from the fact that every function having the form $c \cdot \alpha^n + d$ is convex in n for $c \geq 0$ and $\alpha \geq 0$. Finally, the fact that A_n is increasing in n follows directly from Eq (5.20).

□

Lemma 5.3. *If* $\tau > \frac{-\log(1-p\cdot s)}{\kappa}$*, then then the sequence* $\{A_n\}$ *converges monotonically to* $A = \frac{\beta}{1-\alpha}$*.*

Proof. First, rewriting yields that the condition $\tau > \frac{-\log(1-p\cdot s)}{\kappa}$ is equivalent to the condition $\alpha = \frac{1}{1-p\cdot s} e^{-\kappa\cdot\tau} < 1$. The fact that the sequence converges to A now follows immediately by taking the limit in Equation (5.21):

$$\lim_{n\to\infty} A_n = \lim_{n\to\infty} \alpha^n \cdot A_0 + \beta \frac{\alpha^n - 1}{\alpha - 1} \tag{5.24}$$

$$= \frac{\beta}{1-\alpha} = A \tag{5.25}$$

To prove that A_n converges *monotonically* to A, we distinguishing two cases: $A_{n-1} < A$ and $A_{n-1} > A$, (it is obvious for $A_{n-1} = A$). First, consider the case that $A_{n-1} < A$. We show that $A_n < A$, and that $A_n - A_{n-1} > 0$, implying monotonicity. The first claim is implied by Equation (5.20):

$$A_n = \alpha \cdot A_{n-1} + \beta \tag{5.26}$$

$$< \alpha \cdot \frac{\beta}{1-\alpha} + \beta \tag{5.27}$$

$$= \frac{\beta}{1-\alpha} = A \tag{5.28}$$

To prove the second claim we make use of the given that $\alpha < 1$:

$$A_n - A_{n-1} = \alpha \cdot A_{n-1} + \beta - A_{n-1} \tag{5.29}$$

$$= (\alpha - 1)\, A_{n-1} + \beta \tag{5.30}$$

$$> (\alpha - 1)\, \frac{\beta}{1-\alpha} + \beta = 0 \tag{5.31}$$

Next, consider the case that $A_{n-1} > A$. Using on exactly the same reasoning as the one presented above, it follows that $A_n > A$ as well, and that $A_n - A_{n-1} < 0$.

This completes the proof. □

Now we have established the (long term) behavior of the parameter A_n, let us consider the implications for the (long term) prevalence level. Recall that $\bar{f}_{n,n+1}$ denotes the expected average prevalence level faced between screening rounds n and $n+1$, which is calculated as:

$$\bar{f}_{n,n+1} = \frac{1}{\tau}\int_0^\tau \frac{K}{1 + A_n \cdot e^{-\kappa \cdot t}} dt \tag{5.32}$$

$$= \frac{K}{\kappa \cdot \tau} \log\left(\frac{A_n + e^{\kappa \cdot \tau}}{A_n + 1}\right) \tag{5.33}$$

Using this definition, we now prove the desired results:

Proposition 5.1. If $\tau \leq \frac{-\log(1-p\cdot s)}{\kappa}$, then $\lim_{n\to\infty} \bar{f}_{n,n+1} = 0$.

Proof. From Lemma 5.2 we know that $\lim_{n\to\infty} A_n = \infty$. Hence, we derive that:

$$\lim_{n\to\infty} \bar{f}_{n,n+1} = \lim_{n\to\infty} \frac{K}{\kappa \cdot \tau} \log\left(\frac{A_n + e^{\kappa \cdot \tau}}{A_n + 1}\right) \tag{5.34}$$

$$= \frac{K}{\kappa \cdot \tau} \log(1) = 0 \tag{5.35}$$

□

Proposition 5.2. If $\tau > \frac{-\log(1-p\cdot s)}{\kappa}$, then $\lim_{n\to\infty} \bar{f}_{n,n+1} = K\left(\frac{\log(1-p\cdot s)}{\kappa\cdot\tau} + 1\right)$.

Proof. From Lemma 5.3 we know that $\lim_{n\to\infty} A_n = A = \frac{\beta}{1-\alpha}$. Hence, we derive that:

$$\lim_{n\to\infty} \bar{f}_{n,n+1} = \lim_{n\to\infty} \frac{K}{\kappa \cdot \tau} \log\left(\frac{A_n + e^{\kappa \cdot \tau}}{A_n + 1}\right) \tag{5.36}$$

$$= \frac{K}{\kappa \cdot \tau} \log\left((1 - p \cdot s)\, e^{\kappa \cdot \tau}\right) \tag{5.37}$$

$$= K\left(\frac{\log(1 - p \cdot s)}{\kappa \cdot \tau} + 1\right) \tag{5.38}$$

□

5.C Proofs on Screening Requirements for Elimination

For sake of conciseness, we omit the subscript v in this section. Recall that $\alpha = \frac{1}{1-p\cdot s}e^{-\kappa\cdot\tau}$, $\beta = \frac{p\cdot s}{1-p\cdot s}$, that $A_0 = \frac{K}{f(0)} - 1$, and that α, β, and A_0 are strictly positive.

Proposition 5.3. If $f(0) > C$ and $\tau \leq \frac{-\log(1-p\cdot s)}{\kappa}$, the expected prevalence level is smaller than or equal to C after screening round n^*, where

$$n^* = \begin{cases} \left\lceil \alpha \log\left(\frac{\frac{K}{C} - 1 + \frac{\beta}{\alpha-1}}{A_0 + \frac{\beta}{\alpha-1}} \right) \right\rceil & \text{if } \alpha > 1 \\ \left\lceil \frac{\frac{K}{C} - 1 - A_0}{\beta} \right\rceil & \text{if } \alpha = 1 \end{cases} \tag{5.39}$$

Proof. First, rewriting yields that the condition $\tau \leq \frac{-\log(1-p\cdot s)}{\kappa}$ is equivalent to the condition $\alpha = \frac{1}{1-p\cdot s}e^{-\kappa\cdot\tau} \geq 1$. According to the LMCCC model, the first time the expected prevalence level crosses the boundary value C occurs immediately after a screening round. This is because the expected prevalence level only increases in the period between two screening rounds. Hence, we prove our theorem by determining the screening round $n^* : f(S^+_{n^*}) \leq C$. By the definition of $A_n = \frac{K}{f(S^+_n)} - 1$, we know that this is equivalent to determining $n^* : A_{n^*} \geq \frac{K}{C} - 1$. Now, substituting the following recurrence relation (see S2 Text)

$$A_n = \begin{cases} \alpha^n \cdot A_0 + \beta \frac{\alpha^n - 1}{\alpha - 1} & \text{if } \alpha > 1 \\ A_0 + n \cdot \beta & \text{if } \alpha = 1 \end{cases} \tag{5.40}$$

yields:

$$\alpha^n \cdot A_0 + \beta \frac{\alpha^n - 1}{\alpha - 1} \geq \frac{K}{C} - 1 \qquad \text{if } \alpha > 1 \tag{5.41}$$

$$A_0 + n \cdot \beta \geq \frac{K}{C} - 1 \qquad \text{if } \alpha = 1 \tag{5.42}$$

Rewriting these inequalities gives the desired result.

□

Lemma 5.1. Given that $\tau = \frac{T}{n} \leq \frac{-\log(1-p\cdot s)}{\kappa}$, then $\lim_{n\to\infty} A_n = \infty$ and the sequence $\{A_n\}$ is monotonically increasing in n.

Proof. To reflect that the value of α now depends on n, the number of screening rounds performed in the next T years, let us define α_n as:

$$\alpha_n = \frac{1}{1 - p \cdot s} \cdot e^{-\kappa \frac{T}{n}} \tag{5.43}$$

Hence, rewriting the definition of A_n in terms of α_n yields:

$$A_n = \alpha_n^n \cdot A_0 + \sum_{i=1}^{n} \alpha_n^{n-i} \cdot \beta \tag{5.44}$$

In what follows, we make use of the fact that $\alpha_n \geq 1$, that $\alpha_n > \alpha_{n-1} > 0$ and hence that $\alpha_n^i > \alpha_{n-1}^i$. This property follows directly from the definition of α_n and from the fact that $\kappa \cdot T$ is strictly positive. Using this result, the first part of the Lemma is shown by taking the limit for n:

$$\lim_{n\to\infty} A_n = \lim_{n\to\infty} \alpha_n^n \cdot A_0 + \sum_{i=1}^{n} \alpha_n^{n-i} \cdot \beta = \infty \tag{5.45}$$

The following inequality proves that $A_n - A_{n-1} > 0$, and hence that A_n increases monotonically with n:

$$A_n - A_{n-1} = A_0 \left(\alpha_n^n - \alpha_{n-1}^{n-1}\right) + \beta \left(\alpha_n^n + \sum_{i=1}^{n-2} \left(\alpha_n^i - \alpha_{n-1}^i\right)\right) > 0 \qquad (5.46)$$

□

5.D Results for Maximum Likelihood Approaches

	Poisson $ME/MAE/MRE$	Beta-Binomial $ME/MAE/MRE$	Negative Binomial $ME/MAE/MRE$
Model 1: LM	*/*/*	*/*/*	*/*/*
Model 2: FEM	-0.00080/0.00598/1.88	0.00168/0.00702/2.21	-0.000232/0.00523/1.65
Model 3: LMCCC	0.00132/0.00667/2.10	0.01056/0.01537/4.84	0.00689/0.01187/3.74
Model 4: rLMCCC	0.00575/0.01080/3.39	0.03352/0.03802/11.97	0.02248/0.02708/8.52
Model 5: LMVCC	-0.00007/0.00434/1.36	0.00639/0.00943/2.97	0.00194/0.00720/2.27

* Model could not be fitted: several observations require the calculation of $P(x_v(t) = y), y > 0$, whereas $f_v(t) = 0$. Since this probability equals 0 irrespective of the parameter estimates, the (log)likelihood could not be maximized.

Table 5.3: Predictive performance of the five models in terms of mean errors (ME), mean absolute errors (MAE), and mean relative errors (MRE) when they are fitted using a maximum likelihood approach assuming a Poisson, Beta-Binomial, or Negative Binomial distribution.

5.E Sensitivity Analysis on Sensitivity and Participation Levels

	$\rho = -0.2$ $ME/MAE/MRE$	$\rho = -0.1$ $ME/MAE/MRE$	$\rho = 0.05$ $ME/MAE/MRE$
Model 3: LMCCC	$-$**0.00018**/0.00530/1.67	$-$**0.00033**/0.00521/1.64	$-$0.00112/0.00463/1.46
Model 4: rLMCCC	0.00221/0.00728/2.29	0.00150/0.00674/2.12	**0.00014**/0.00577/1.82
Model 5: LMVCC	$-$0.00113/**0.00395/1.24**	$-$0.00136/**0.00382/1.20**	$-$0.00191/**0.00357/1.13**

Table 5.4: Predictive performance of the logistic models in terms of mean errors (ME), mean absolute errors (MAE), and mean relative errors (MRE) when the real case detection fraction ($s \cdot p_{vn}$) deviates a fraction ρ from the values used in the baseline analysis. E.g., $\rho = -0.1$ means that the real fraction is 10% smaller than the presently used fraction. The best indicator values are in bold.

5.F Sensitivity Analysis on the Initial Value Assumption

	$\iota = 0.4$ $ME/MAE/MRE$	$\iota = 0.6$ $ME/MAE/MRE$	$\iota = 0.8$ $ME/MAE/MRE$
Model 3: LMCCC	−0.00086/0.00481/1.51	−0.00095/0.00475/1.50	−0.00096/0.00475/1.49
Model 4: rLMCCC	**0.00069**/0.00619/1.95	**0.00028**/0.00583/1.84	**0.00009**/0.00567/1.79
Model 5: LMVCC	−0.00171/**0.00395**/**1.16**	−0.00186/**0.00369**/**1.12**	−0.00194/**0.00350**/**1.10**

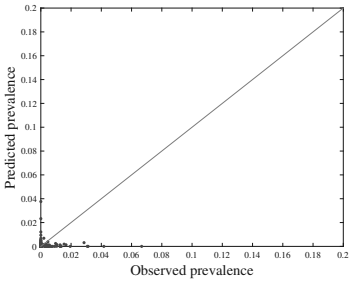
Table 5.5: Predictive performance of the logistic models in terms of mean errors (ME), mean absolute errors (MAE), and mean relative errors (MRE) for different assumptions about ι, the parameter determining the expected prevalence level at the beginning of the time horizon: $f_v(0) = \iota\tilde{x}_v$.

5.G Table of Notations

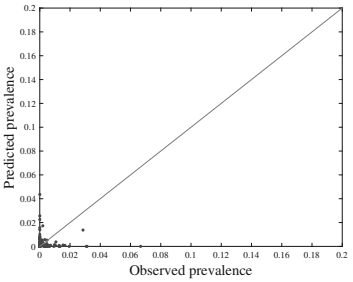
Village characteristics	
V	set of villages v
N_v	population size of village v
p_{vn}	participation fraction in the n^{th} screening round for village v
S_{vn}	time at which the n^{th} screening round in village v is performed
S_{vn}^{+}	moment immediately after the n^{th} screening round in village v
S_{vn}^{-}	moment just before the n^{th} screening round in village v
s_{vn}	time between the $n-1^{th}$ and the n^{th} screening round in village v
$\delta_v^{-}(t)$	time since the last screening round in village v before time t
$\bar{\mu}_v$	average screening frequency in village v during 5 consecutive years
$\bar{\mu}_v(S_{vn})$	average screening frequency in the 3 years prior to the n^{th} screening round in village v
Prevalence level notations	
$x_v(t)$	prevalence level (fraction) in village v observed at time t
$f_v(t)$	expected prevalence level (fraction) in village v at time t
ε_v	random disturbance for the prevalence level (fraction) in village v
$\bar{x}_v$	average observed prevalence level (fraction) in village v during 5 consecutive years
$\bar{x}_v(S_{vn})$	average observed prevalence level (fraction) in the 3 years prior to the n^{th} screening round in village v
Prediction model notations	
α_v	fixed effect for village v
K_v	carrying capacity of village v
κ	growth rate parameter
A_{vn}	initial value parameter for screening round n in village v
e_{vn}	prediction error/ fitting deviation for screening round n in village v
w_{vn}	weight of observation n for village v
Disease characteristics	
r	yearly removal rate for HAT
R_0	basic reproduction number for HAT
s	sensitivity of the diagnostic test
Screening policy analysis notations	
τ	fixed time interval for screening rounds
$\bar{f}_{n,n+1}$	expected average prevalence level between the n^{th} and the $(n+1)^{th}$ screening round

Table 5.6: Table of Notations.

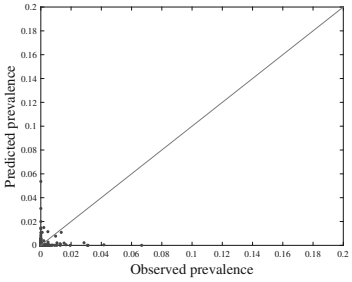
5.H Prediction Errors



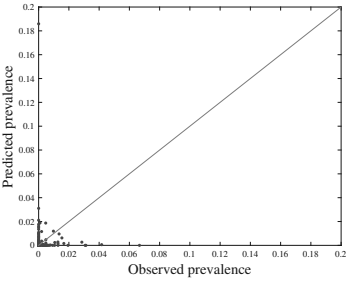
(a) LM model



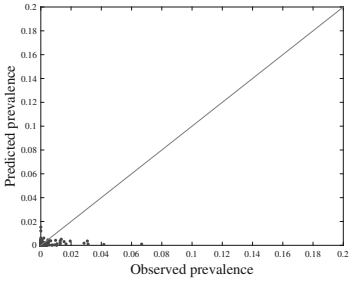
(b) FE model



(c) LMCCC model



(d) rLMCCC model



(e) LMVCC model

Figure 5.7: Predicted prevalence vs. observed prevalence for the 143 observations in the prediction sample.

Chapter 6

Optimizing Population Screening for Infectious Diseases

6.1 Introduction

Infectious diseases are disorders caused by the entrance, growth, and multiplication of microorganisms, such as bacteria, viruses, parasites or fungi, in the body (Bennett et al., 2014). Despite optimism about the elimination of these diseases in the 1960s, fuelled by improved sanitation, antibiotics, and vaccination programs, they continue to be a major cause of suffering. Infectious disease agents have evolved, new diseases have emerged, some have re-emerged, and for others considerable efforts have not led to elimination (Hethcote, 2000). Some of the most notable examples are HIV, tuberculosis (TB), and malaria.

An important element in fighting infectious diseases is to detect and control outbreaks as early as possible (Dasaklis et al., 2012). This is particularly important for high-impact diseases that have a high transmission rate and/or have a long incubation period. The success of disease control programs therefore often depends on

active case finding (ACF), i.e. the deployment of health service staff to search for infected individuals. ACF forms an important practice in the fight against disease such as tuberculosis (Golub et al., 2005), leprosy (Moura et al., 2013), and human African trypanosomiasis (HAT) (WHO, 2013). One of the most rigorous ACF strategies is to perform exhaustive population screening – also referred to as door-to-door screening – in which mobile teams are sent out to screen villages or populations at risk.

Though population screening is costly, it can effectively contain epidemics and can be a crucial element of the journey towards disease elimination (cf. WHO, 2013). Yet, the willingness of health authorities and international donors to maintain this costly strategy tends to diminish when the urgency, e.g. in terms of number of cases found, decreases (see Hasker et al., 2012, 2010, for evidence for HAT). As a result, the disease persists and incidence and prevalence may increase again, thus increasing the burden of disease. To improve the applicability and acceptability of this tool, it is hence of key importance to deploy the mobile teams so as to maximize the *effectiveness* of the population screening. The latter is to a large extent determined by the planning decisions to determine *which villages* to screen and with *what time interval* to screen them (cf. Mpanya et al., 2012). Yet, research on how to plan screening *optimally* appears to be lacking, as for instance indicated by the WHO in the context of HAT control (WHO, 2015b). In fact, current planning practices are reported to vary widely (see, e.g., Ruiz et al., 2002; Simarro et al., 1990; Paquet et al., 1994).

This chapter investigates how to maximize effectiveness of population screening for infectious diseases. The problem of maximizing effectiveness is formulated as deploying mobile screening teams so as to minimize the expected burden of an infectious disease, which we refer to as the mobile screening team deployment problem ($MSTD$). The novelty and complexity of $MSTD$ lie in the objective function which must capture the effect of planning on the expected burden of disease. This relationship is considered to be complex and highly non-linear (see for instance Chapter 5 and the references therein). Our general model assumes that the expected case load and disease burden decreases after a screening round, due to treatment initiation for the cases identified. Motivated by existing work (cf. Welburn et al., 2016; Robays

et al., 2004; Piot, 1967), we consider two assumptions to capture this effect: (1) the disease burden in the village drops to a fixed level, or (2) the disease burden decreases by a given fraction after screening. For the period *between* two screening rounds, we solely assume the disease burden to be increasing. We propose several flexible solution approaches to solve $MSTD$ as well as simple planning policies, which tend to be more suitable in practice. Moreover, we use insights into the disease burden associated with optimal planning decisions to assess the relationship between screening capacity and long-term disease burden. This can assist policy makers in determining how many teams are expected to be needed to stabilize the epidemic or to reach eradication of the disease.

Our results and methods are generic, and have much potential to be of value in solving pressing real-life problems. This is illustrated through the analysis of a special case of the deployment problem, the problem of planning mobile screening teams for HAT, which motivated our work. This special case is referred to as $MSTD_{HAT}$. Approximately 70 million people in 36 Sub-Saharan countries are at risk of HAT infection (Simarro et al., 2012; WHO, 2015a), and massive population screening has so far been the sole effective control strategy. We show that easily implementable planning policies yield near-optimal decisions and have the potential to substantially improve upon the policy recommended by the WHO. Furthermore, we prove a piecewise linear relationship between expected HAT prevalence and screening capacity.

Below, we review related literature, formally define $MSTD$, and present basic models for the expected burden of disease. We then present generic methods to assess the relationship between screening capacity and the expected disease burden. We develop three general solution approaches and four simple planning policies for $MSTD$ and show $MSTD$ to be NP-Hard. Finally, we numerically assess the performance of the algorithms and the robustness of the solutions for $MSTD_{HAT}$. All proofs can be found in Appendix 6.A.

6.2 Literature

Optimization in Epidemic Control. Logistics plays an essential role in epidemic control. It determines the speed of response, which in turn determines the disease

burden averted. Dasaklis et al. (2012) provide an overview of scientific research in this field. Applications include network design for medical supplies, inventory management, and triage strategies. The authors observe a substantial gap in literature explicitly incorporating logistics operations: Most research has focused on strategic resource allocation problems.

The latter arise when decision makers have to allocate limited resources among competing control programs or populations. See Katoh and Ibaraki (1998) for an overview of the corresponding literature. The characteristics of an epidemic make these problems complex (Brandeau, 2005). First, epidemics typically evolve in a non-linear way (see Hethcote, 2000). Second, the relationship between the amount of resources invested and health outcomes, which is referred to as the "production function" (Zaric and Brandeau, 2001), tends to be complex. Third, control programs or populations may not be independent.

Our problem has some similarity with resource allocation problems in the sense that screening capacity has to be "allocated" among different villages. Yet, it focuses on operational resource allocation decisions, i.e. the decisions which villages to visit and when, whereas strategic (and tactical) level resource allocation has focused on resource quantities and corresponding static allocation decisions.

Mobile Healthcare Planning Problems. Distance or travel time to the nearest healthcare facility is a major determinant of healthcare utilization and several types of health outcomes (De Vries et al., 2014a). Providing sufficient levels of access through spatially fixed facilities is often not feasible, particularly in scarcely populated and poor areas. For this reason, mobile healthcare units are being used in several countries (see Doerner et al. (2007) for references).

The routing problem for mobile healthcare facilities was first presented by Hodgson et al. (1998). The authors model this as a tour-location problem, which is to select a tour and tour stops so as to minimize the total travel time and to satisfy the constraint that each demand point is covered. Hachicha et al. (2000) extend this model to the multiple vehicle case and propose three heuristics to solve it. Doerner et al. (2007) model the problem as a multi-objective optimization problem, using coverage and travel time criteria. The main difference between these problems and

$MSTD$ is that the latter minimizes the burden of disease, i.e. a health outcome measure, rather than travel times or coverage, which are logistics process measures.

Finally, McCoy and Lee (2014) analyze optimal deployment of motorcycles, which are used to provide healthcare services in rural areas. The problem of determining the number of visits to each outreach site is modeled as a resource allocation problem based on effectiveness and equity criteria. Effectiveness is modeled to depend (polynominally) on the number of visits. In $MSTD$, effectiveness is modeled more specifically by means of disease burden as a function of time intervals between consecutive visits.

Maintenance Planning Problems. $MSTD$ is closely related to maintenance optimization problems. Using the classification scheme presented by Dekker (1996), a village can be modeled as a multi-component system, where the inhabitants are the components. Disease burden can be modeled as deterioration, risk of infecting others as stochastic dependence between components, and screening as maintenance or repair.

Multi-component maintenance models are reviewed by Nicolai and Dekker (2008). A vast majority of these models, however, do not take resource constraints into consideration but instead make a trade-off between costs and benefits of maintenance. Here, we mention papers that most closely resemble our work. Dekker (1995) proposes a MIP model to prioritize a fixed number of maintenance activities. The objective is to minimize the total penalty associated with delays. Wei and Liu (1983) analyze the problem of designing a schedule that adheres to capacity constraints and *fixed* given maintenance time intervals. A close variant of this problem is the free periodic maintenance problem (FPMP), which is to find a cyclic schedule that minimizes the sum of operating costs and maintenance costs (Anily et al., 1998; Bar-Noy et al., 2002). The periodic maintenance problem (PMP) additionally assumes a fixed number of planning periods (Grigoriev et al., 2006). We refer to the latter work for an overview of results on these and similar problems.

$MSTD$ differs from PMP in two ways. First, PMP assumes that at most M machines can be serviced per period, whereas in $MSTD$ capacity is also geographically and demographically constrained. Second, the assumption that disease burden in-

creases linearly with the time without screening and is "reset" to zero after screening will generally not hold for epidemics.

Finally, our problem has some similarity with the k-travelling repairman problem, which is to select tours for k repairmen so as to minimize the total waiting time of the customers (Fakcharoenphol et al., 2003). Again, there are fundamental differences in terms of objective functions and geographical constraints.

Human African Trypanosomiasis. HAT or sleeping sickness is a slowly progressing parasitic disease, transmitted from human to human through the tsetse fly (Brun et al., 2010). The case that we elaborate on in this chapter is the T.B. Gambiense variant of HAT, which causes 98% of all HAT cases (WHO, 2015a). It is fatal if left untreated and develops in two stages. In the first stage, the parasite typically causes little or unspecific symptoms such as headaches, fever, and weakness (Brun et al., 2010). The median duration of this stage is about 1.5 years (Checchi et al., 2008). Patients typically do not seek care until a second stage, when the parasite has crossed the blood-brain barrier. The parasite then causes various neurological disorders, including sleeping disorders, severe suffering, and, eventually, death.

The treatment delay associated with the first asymptomatic phase is a major enabler of sustained transmission of HAT. Patients are a potential source of infection for the tsetse fly (Brun et al., 2010; Fevre et al., 2006), and hence indirectly for uninfected people. They may be infectious for more than 1.5 years until they start to seek care themselves (Hasker et al., 2010); this is considered a form of passive case finding (PCF). These characteristics underline the crucial importance of active case finding (ACF) and early treatment for effective HAT control (WHO, 2013; Hasker et al., 2010).

The current practice of ACF is to send mobile teams to endemic villages for exhaustive population screening (Mpanya et al., 2012; Brun et al., 2010; Hasker et al., 2010). The national HAT disease control program in the DRC, for example, employs 35 such mobile screening teams. This strategy is proven to be effective (Rock et al., 2015; Fevre et al., 2006), but is also considered to be costly. As a consequence, funding and thus screening activity may be reduced (Hasker et al., 2010), resulting in a risk of increased HAT incidence and prevalence (Simarro et al., 2011). Major

epidemics following the collapse of screening after independence of African countries illustrate the need for sustained screening to prevent reemergence (Brun et al., 2010). The WHO, for instance, estimated that more than 300 000 cases occurred in 1998 alone (WHO, 2015a).

HAT has been targeted for elimination by the WHO. Elimination will require considerable effectiveness from the scarcely available resources, thus further underlining the need for optimization. Chapter 5 studies several models to capture the effects of ACF planning decisions on HAT prevalence, which we use as a basis for analysis.

6.3 Problem Formulation

In $MSTD$, we consider M mobile teams, a set of endemic villages $\mathcal{V}$ and planning periods (e.g., months) $t \in \mathcal{T} = \{1, 2, \ldots, T\}$. $MSTD$ entails determining for each mobile team and for each period t which subset of villages to visit, so as to minimize the total expected disease burden over the planning horizon.

The assignment of teams to villages is constrained by a *maximum mission length*. One mission corresponds to one planning period, and its length equals the time on site plus the time spent traveling. Whereas the time on site can be easily estimated, there are several reasons to simplify the estimation of the time spent traveling. First, due to time-consuming screening efforts, time on site tends to dominate travel times. Second, travel time data may be scarce and imprecise. Third, time on site and travel are often combined, causing working days to be a bit longer or shorter depending on the required travel. Instead of considering routing and travel time details, we therefore assume that a team stays within the same region or *cluster* of villages during the planning period for which travel times between subsequently visited villages can be conveniently incorporated in daily schedules. A cluster c in the set of clusters $\mathcal{C}$ hence represents a collection of villages that lie "within travel distance" from each other. We denote the subset of villages corresponding to cluster c by $\mathcal{V}_c$ and let binary variables y_{ct} indicate whether a team is assigned to cluster c in period t.

We assume that screening village v in a given period use up a fraction r_v of the time available in that planning period. We let binary variables x_{vt} indicate whether village v is screened in planning period t. These variables translate into a vector

of time intervals between consecutive screening rounds $\tau_v(\mathbf{x}_v) = \{\tau_{v0}, \tau_{v1}, ..., \tau_{vn_T}\}$. Here, τ_{v0} represents the time between the beginning of the planning horizon and screening round 1, τ_{v1} the time between screening round 1 and screening round 2, ..., and τ_{vn_T} the time between the last screening round and the end of the planning horizon. For convenience of notation, we will denote this vector by τ_v from now on. We let function $\mathcal{B}_v(\tau_v)$ represent the resulting average expected disease burden over the planning horizon. Disease burden could, for example, be measured in terms of average prevalence, person-years of illness, or QALYs or DALYs (Yadav, 2010). Using this notation, $MSTD$ can be formulated as:

$$\min \sum_{v \in \mathcal{V}} \mathcal{B}_v(\tau_v) \tag{6.1}$$

$$\text{s.t.} \sum_{v \in \mathcal{V}_c} r_v x_{vt} \leq y_{ct} \qquad c \in \mathcal{C}, t \in \mathcal{T} \tag{6.2}$$

$$\sum_{c \in \mathcal{C}} y_{ct} = M \qquad t \in \mathcal{T} \tag{6.3}$$

$$x_{vt}, y_{ct} \in \{0, 1\} \qquad v \in \mathcal{V}, c \in \mathcal{C}, t \in \mathcal{T} \tag{6.4}$$

For each period t and cluster c, Constraint 6.3 regulates screening capacity available. Constraint 6.4 limits the number of teams assigned to clusters per period. The next section discusses the objective function in detail based on the following assumptions:

Assumption 6.1. *The time interval between consecutive screening rounds in periods t and $t + \tau$ equals exactly τ periods. Hence, the precise screening moment within the period is the same for each period.*

Assumption 6.2. *The expected disease burden is increasing in time since the last screening round.*

The first assumption is justifiable when the precise timing of the screening round within the period has little impact on the expected development of the disease burden, which is particularly the case for slowly evolving epidemics. For ease of exposition, we will assume that screening rounds take place *at the end of the planning*

period. The second assumption is very general, and holds for progressive infectious disease for which cure without treatment is not possible. Reduction from the periodic maintenance problem implies the following (see Appendix 6.A):

Proposition 6.1. *The $MSTD$, as formulated in* (6.1) - (6.4)*, is NP-Hard under Assumption 6.2.*

We briefly note here that the decision version of $MSTD$ is not trivially in NP, as it may be non-trivial to evaluate the objective function (or any other certificate for a *yes*-instance) in time that is polynomial in the input parameters.

6.4 Disease Burden Models

For every village $v \in V$, we define the *disease burden progression function* $f_v(s) \geq 0$ to describe the development of the expected disease burden over time from the very beginning of the epidemic in the absence of ACF., i.e. $x_{vt} = 0 \quad \forall t$. Here, s is referred to as the *stage of progression* of the epidemic. The effect of a screening round at stage s is now modeled to reset the epidemic to a different stage of progression. Let s_{vn} denote the stage of progression after screening round n. In the following paragraphs, we consider two cases. In the first case, the disease burden decreases with a given positive fraction p, and the epidemic moves to the stage corresponding to the resulting burden:

$$f_v(s_{vn}) = (1-p) f_v(s_{vn-1} + \tau_{vn-1}). \tag{6.5}$$

This may occur, for example, when only a fraction of the infected people participate in screening rounds or when the diagnostic test only detects a fraction of the cases. In the second case, the epidemic is always reset to a same stage s_v^* after screening. For example, this may occur when *all* infected people are detected during a screening round or when a reservoir of "latent disease carriers" exists who are not detected during screening rounds (see, e.g., Welburn et al., 2016). We refer to these as the *fractional screening impact* case and the *fixed screening impact* case, respectively.

The first is further explained in Figure 6.1. The black line in this figure represents the disease burden progression function. After screening round n, we are at the stage corresponding to the left red point, after which the burden develops to the right red point. Next, screening round $n+1$ makes the burden decrease with fraction p, causing us to end up in an earlier epidemic stage: the one corresponding to the left blue point. Afterwards, the burden develops to the right blue point, and the process repeats.

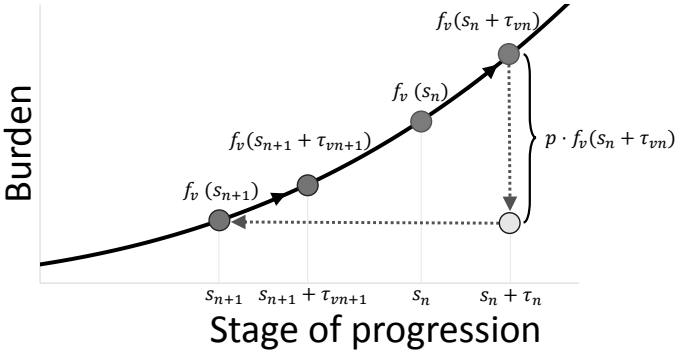


Figure 6.1: Function $f_v(s)$, the disease burden progression function, and the relationship with the actual burden incurred over time for the fractional screening impact case.

We denote the *cumulative disease burden progression function* by:

$$F_v(\tau) = \int_0^{\tau} f_v(s)ds. \tag{6.6}$$

To summarize, we are interested in the average expected disease burden, which relates to decision variables τ_v and progression function $f_v(s)$ as follows:

$$\mathcal{B}_v(\tau_v) = \frac{1}{T}\sum_{n=0}^{n_T}\int_0^{\tau_{vn}} f_v(s_{vn}+t)dt. \tag{6.7}$$

Increasing disease burden progression functions. We first discuss the general class of epidemics for which the burden progression function is simply increasing in time and afterwards consider the special case of a burden progression function for the HAT disease, as proposed in the literature. A property that will be important in our further analyses is that the *cumulative* progression function $F_v(\tau)$ of an increasing progression function, as defined in (6.6), is convex in τ (Hardy et al., 1952).

One general and relevant subclass of increasing burden progression functions is obtained in that case when (1) expected incidence (rate of new disease cases) is a differentiable, increasing function $\lambda_v(s) \geq 0$ of s and/or (2) the expected disease burden incurred by an infected person is a differentiable, increasing function $\beta(\tau - s) \geq 0$ of the time since the onset of the disease $\tau - s$. The disease burden at stage τ is then calculated as:

$$f_v(\tau) = \int_0^\tau \lambda_v(s)\beta(\tau - s)ds, \tag{6.8}$$

and is stated to be increasing in the following proposition.

Proposition 6.2. *The expected disease burden at stage τ, as defined in* (6.8)*, is increasing in τ if functions $\lambda_v(s) \geq 0$ and/or $\beta(s) \geq 0$ are differentiable and increasing in s.*

Logistic model. The average expected number of infected people is a common indicator of disease burden. In an extensive modeling study, De Vries et al. (2016) investigate how expected HAT prevalence, defined as the *fraction* of people infected, in a village v relates to the timing of screening rounds in that village. Based on predictive performance and theoretical justification, variants of the so-called *logistic model* are judged to be the most suitable.

As can be seen from Figure 6.2, the logistic model corresponds to an epidemic in which expected prevalence grows exponentially in an initial phase and levels off to an equilibrium prevalence level or *carrying capacity* afterwards. The latter is village-specific and can be estimated based on past prevalence levels and screening rounds. Let K_v denote the carrying capacity of village v. Then the model specifies that, in

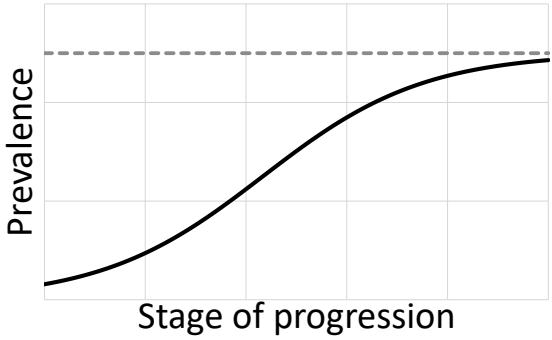


Figure 6.2: Development of expected disease prevalence according to the logistic model. Prevalence first grows exponentially and levels off to the equilibrium prevalence level (dotted line) afterwards.

the absence of screening, the expected number of people infected in village v, having total population N_v, develops according to the following formula:

$$f_v(s_{vn} + t) = \frac{N_v K_v}{1 + Ae^{-\kappa(s_{vn}+t)}}. \tag{6.9}$$

Here, κ represents a constant determining the steepness of the s-shaped curve and $A_v = \frac{K_v}{f_v(0)} - 1$ reflects the prevalence level at the beginning of the epidemic. For convenience in notation, we define $A_{vn} = A_v e^{-\kappa s_{vn}}$. Substituting function (6.9) into (6.7) and deriving the integral yields the following expression for the average expected number of people infected (Chapter 5):

$$\mathcal{B}_v(\tau_v) = \frac{N_v K_v}{\kappa T} \sum_{n=0}^{n_T} \log\left(\frac{A_{vn} + e^{\kappa \tau_{vn}}}{A_{vn} + 1}\right). \tag{6.10}$$

We note that variable $A_{vn}, n > 0$ equals ∞ when $p = 1$ and that it can be determined recursively when $p > 0$ (Chapter 5):

$$A_{vn+1} = \frac{e^{-\kappa\tau_{vn}}}{(1-p)} A_{vn} + \frac{p}{1-p}. \tag{6.11}$$

Let us now turn our attention to the case motivated before where a screening round resets expected prevalence to a fixed level. Letting parameter $A_v^* = e^{-\kappa s_v^*}$ reflect this fixed level, equation (6.10) implies that the average expected number of people infected in village v can be estimated as:

$$\mathcal{B}_v(\tau_v) = \frac{N_v K_v}{\kappa T} \sum_{n=0}^{n_T} \log\left(\frac{A_v^* + e^{\kappa\tau_{vn}}}{A_v^* + 1}\right). \tag{6.12}$$

We refer to the $MSTD$ using the HAT burden function (6.10) or (6.12) as $MSTD_{HAT}$ and prove the following result by reduction from the 3-partition problem:

Proposition 6.3. *$MSTD_{HAT}$ is strongly NP-Hard, even when $M = |\mathcal{C}| = 1$.*

6.5 Relationship between Capacity and Burden

As motivated, insight into the relationship between screening capacity – i.e., the number of mobile screening teams – and disease burden is crucial to assess what resources are needed to eradicate the disease or to keep the disease burden below certain limits in the long term. Of course, one could analyze this by simply solving $MSTD$ for varying capacity levels. However, given the complexity of $MSTD$ and the long-term horizon that needs to be considered, doing so will generally be computationally infeasible. As an alternative, we propose simple methods to assess long-term burden for a stylized variant of $MSTD$, named $MSTD_R$. In this problem, we relax the requirements that at most M teams can be deployed in each time period t and that teams cannot visit multiple clusters in one period. Instead, $MSTD_R$ requires that at most M teams are deployed *on average* per planning period and that each village has an infinite sequence of screening intervals of equal length, denoted by parameter τ_v, or equivalently maintains a fixed screening frequency π_v over an infinite horizon:

$$\min \sum_{v \in \mathcal{V}} \mathcal{B}_v \left(\frac{1}{\pi_v} \right) \tag{6.13}$$

$$\text{s.t.} \sum_{v \in \mathcal{V}} \pi_v r_v \leq M \tag{6.14}$$

$$\pi_v \geq 0 \qquad v \in \mathcal{V}. \tag{6.15}$$

Note that we now define $\mathcal{B}_v$ as a function of fixed screening interval τ_v (instead of a vector of screening intervals). Note further that $MSTD_R$ makes clustering irrelevant. The impact of these assumptions on the implied relationship between capacity and burden is discussed in the final section.

Optimal resource allocation for increasing burden progression functions and fractional screening impact. Solving $MSTD_R$ for the fractional impact case requires insight into the long-term development of the disease burden. It is intuitively clear that, for a given screening interval, the disease burden will either go to infinity in the long term or converge to some value, as formally proven in the following lemma.

Lemma 6.1. *For a given value τ_v and for each increasing progression function $f_v(\tau_v)$, the sequence $\{f_v(s_{vn})\}_n$ either monotonically converges to some finite value or goes to infinity when $n \to \infty$.*

For a given increasing burden function $f_v(s)$, multiple or even infinitely many limiting values of $\{f_v(s_{vn})\}_n$ may exist. The specific value that will be attained is determined by the initial disease burden $f_v(s_{v0})$ and the screening interval τ_v, as shown in Example 6.1. The limiting values can be found by solving:

$$f_v(s) = (1-p) f_v(s + \tau_v) \qquad t \geq 0. \tag{6.16}$$

Let us consider a village v with initial burden $f_v(s_{v0})$ and let Λ denote the (possibly infinite) set of limiting values $f_v(s_\ell)$ that can be attained by varying τ_v. Arguments

similar to those in Lemma 6.1 show that decreasing the screening interval never yields a higher limiting value. As a consequence, it is not possible that screening intervals τ_1 and τ_2 lead to the same limiting value whereas τ_3, $\tau_1 < \tau_3 < \tau_2$ leads to a different limiting value. Hence, there exists a partition of values for τ_v in a (possibly infinite) set of domains, e.g. $(\underline{\tau}_{v\ell}, \overline{\tau}_{v\ell}]$, such that each domain is associated with a single limiting value. Within the domain, it holds that setting $\tau_v = \overline{\tau}_{v\ell}$ *dominates* other choices in the sense that it yields the same long-term average disease burden $\mathcal{B}_{v\ell} = \frac{1}{\overline{\tau}_{v\ell}} \int_0^{\overline{\tau}_{v\ell}} f_v(s_\ell + t)dt$ with the lowest screening frequency. This follows from the fact that average disease burden, as defined in (6.7), is calculated as the Cesaro mean of a converging series, which is shown to equal the limiting value of that series (Hardy, 2000).

If the number of limiting values is finite, the problem is to select one of the finite set of dominating screening frequencies for each village such that we minimize average expected disease burden and obey the capacity constraint. Hence, problem (6.13) - (6.15) can be solved as a multiple-choice knapsack problem (Kellerer et al., 2004) with capacity M, weights $w_{v\ell} = \lim_{\tau_v \to \frac{r_v}{\overline{\tau}_{v\ell}}}$ and values $\mathcal{B}_{v\ell}$. This method is not feasible if there is an infinite number of limiting values. For several functions, however, the limiting value can be expressed as a function of τ_v and the resulting optimization problem can be efficiently solved, as illustrated in Example 6.2.

Example 6.1. Suppose that $f_v(s) = e^{\alpha_v s+\beta_v}$. Then it holds that:

$$\begin{aligned}\frac{f_v(s_{vn+1})}{f_v(s_{vn})} &= \frac{(1-p)e^{\alpha_v(s_{vn}+\tau)+\beta_v}}{e^{\alpha_v s_{vn}+\beta_v}} \\ &= (1-p)e^{\alpha_v \tau_v}.\end{aligned}$$

Hence there exists some threshold $\tau_v^* = \frac{1}{\pi_v^*}\log\left(\frac{1}{1-p}\right)/\alpha_v$ such that:

$$\lim_{n\to\infty} f_v(s_{vn}) = \begin{cases}\infty & \text{if } \tau_v > \tau_v^* \\ f_v(s_{v0}) & \text{if } \tau_v = \tau_v^* \\ 0 & \text{if } \tau_v < \tau_v^*\end{cases}$$

This shows that, under optimal resource allocation, there exists a threshold capacity $M^* = \sum_{v\in\mathcal{V}} \pi_v^* r_v$ such that the long-term average disease burden equals 0 if M exceeds this capacity, goes to infinity if M is smaller than this capacity, and equals $\sum_{v\in\mathcal{V}} f_v(s_{v0})$ otherwise.

Example 6.2. Suppose that $f_v(s) = \alpha_v s^2$. Then equation (6.16) yields:

$$\frac{\alpha_v(s+\tau_v)^2}{\alpha_v s^2} - \frac{1}{1-p} = 0.$$

Calculating the roots of this quadratic function gives the following limiting progression stage:

$$s_v^* = \frac{2\tau_v + \sqrt{p/(1-p)}}{\sqrt{p/(1-p)}}.$$

Plugging this into $f_v(s)$ yields the following limiting value for the disease burden:

$$\lim_{n\to\infty} f_v(s_{vn}) = \alpha_v\left(\frac{2\tau_v + \sqrt{p/(1-p)}}{\sqrt{p/(1-p)}}\right)^2$$

Hence, $f_v(s_{vn})$ will converge to $f_v(s_v^*)$, where s_v^* is a linear function in τ_v. After applying variable transformation $\pi_v = \frac{1}{\tau_v}$, the resulting allocation problem can be solved as a convex optimization problem.

Optimal resource allocation for increasing burden progression functions and fixed screening impact. In the fixed screening impact case, the disease burden incurred between two screening rounds remains the same for each pair of consecutive rounds. For this reason, one only needs to analyze the burden between *one pair of screening rounds* rather than its long-term development. Next, we propose a general and practically suitable solution method for problem (6.13) - (6.15) based on the following lemma:

Lemma 6.2. *Let $f_v(\tau_v)$ be an increasing burden progression function. Then $\mathcal{B}_v\left(\frac{1}{\pi_v}\right)$ is convex and decreasing in π_v for $\pi_v > 0$.*

This implies that, for all increasing burden progression functions $f_v(\tau_v)$, the problem can be solved efficiently as a convex optimization problem, e.g. using interior point algorithms (Nesterov and Nemirovskii, 1994). These methods also effectively deal with the boundary issue that arises when $F_v\left(\frac{1}{\pi_v}\right)$ is not defined for $\pi_v = 0$ but only for $\lim \pi_v \to 0$, as they approach but never actually reach the boundary (Nesterov and Nemirovskii, 1994).

Optimal resource allocation for the logistic model with fractional screening impact. Since the disease burden progression function corresponding to the logistic model is increasing, the results presented above also hold for this specific case. Dedicated analysis of the optimal resource allocation, however, yields more specific insights.

Let $\bar{f}_{vn}$ denote the average expected prevalence in village v between screening rounds n and $n+1$. The following lemma states how the long-term average expected disease burden relates to τ_v.

Lemma 6.3. *Screening village v with constant interval τ_v yields average prevalence level:*

$$\begin{aligned}\mathcal{B}_v(\tau_v) &= \lim_{n_T \to \infty} \frac{1}{n_T} \sum_{n=0}^{n_T} N_v \bar{f}_{vn} \\ &= \max\left\{0, N_v K_v \left(\frac{\log(1-p)}{\kappa \tau_v} + 1\right)\right\}. \end{aligned} \tag{6.17}$$

The following lemma implies that we can get rid of the maximum operator:

Lemma 6.4. *Under the logistic model with fractional screening impact, there exists an optimal solution to problem* (6.13)-(6.15) *in which the screening interval is at least $\tau^* = \frac{-\log(1-p)}{\kappa} > 0$ for each village.*

Enforcing this lower bound implies that the second term in (6.17) is nonnegative. Consequently, the burden function can be replaced by this second term itself. This transforms problem (6.13) - (6.15) into the following LP problem:

$$\min \sum_{v \in \mathcal{V}} N_v K_v \left(\pi_v \frac{\log(1-p)}{\kappa} + 1 \right) \tag{6.18}$$

$$\text{s.t.} \sum_{v \in V} \pi_v r_v \leq M \tag{6.19}$$

$$\pi^* \geq \pi_v \geq 0 \qquad v \in \mathcal{V} \tag{6.20}$$

Here, $\pi^* = \frac{1}{\tau^*}$. The following proposition states the optimal solution.

Proposition 6.4. *Under the logistic model with fractional screening impact, the total average expected disease burden is minimized by greedily assigning screening interval*

$$\tau = \max \left\{ \left(-\log\left(1-p\right) \right) / \kappa, \tau_R \right\}$$

to villages in descending order of the ratio $N_v K_v / r_v$. Here, τ_R denotes the minimum screening interval that can feasibly be attained using the remaining screening capacity.

Figure 6.3 illustrates this result. For village v, increasing screening frequency linearly from zero to $r_v \pi^*$ decreases the burden linearly from $N_v K_v$ to zero. Hence, doing so in descending order of the presented ratio yields a piecewise linear relationship between capacity M and total average expected disease burden. Note that it is easy to read from this figure the minimum capacity required for reaching a given target burden level.

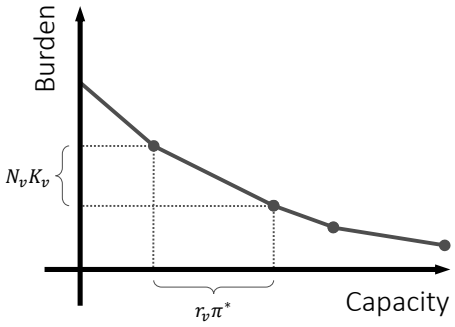


Figure 6.3: Piecewise linear relationship between capacity M and average expected disease burden, as implied by the logistic model with fractional screening impact.

Optimal resource allocation for the logistic model with fixed screening impact. Let us now consider the optimal resource allocation when the disease burden develops according to the logistic model with absolute screening impact. Screening village v at frequency π_v indefinitely yields the following average expected disease burden (see equation (6.12)):

$$\mathcal{B}_v\left(\frac{1}{\pi_v}\right) = \frac{N_v K_v \pi_v}{\kappa} \log\left(\frac{A_v^* + e^{\frac{\kappa}{\pi_v}}}{A_v^* + 1}\right). \tag{6.21}$$

Let us define $\mathcal{B}_v\left(\frac{1}{0}\right)$ as the average disease burden in village v when it is never visited by a screening team. i.e., $\mathcal{B}_v\left(\frac{1}{0}\right) = \lim_{\pi_v \downarrow 0} \mathcal{B}_v\left(\frac{1}{\pi_v}\right)$. In Appendix 6.A we prove the following, which is in line with the general results proven in Lemma 6.2:

Lemma 6.5. *The disease burden function* $\mathcal{B}_v\left(\frac{1}{\pi_v}\right)$*, as defined in* (6.21)*, is convex and decreasing in* π_v *for* $\pi_v \geq 0$*.*

Because of the convexity of the objective function and the linearity of the resource constraint, problem (6.13) - (6.15) is equivalent to the following Lagrangean dual optimization problem, which can be solved using a subgradient algorithm (Bertsekas, 1999):

$$\max\{\mathcal{L}(\lambda) : \lambda \leq 0\}. \tag{6.22}$$

Here, the Lagrangean function $\mathcal{L}(\lambda)$ is defined as:

$$\mathcal{L}(\lambda) = \min\left\{\sum_{v\in V} \mathcal{B}_v\left(\frac{1}{\pi_v}\right) + \lambda\left(M - \sum_{v\in V}\pi_v r_v\right) : \pi_v \geq 0\right\}. \tag{6.23}$$

The following result provides the optimal solution to the Lagrangean function. Here, function $\left(\frac{d}{d\tau}\mathcal{B}_v\right)^{-1}$ denotes the inverse of the derivative of $\mathcal{B}_v$:

Proposition 6.5. *The optimal value of $\mathcal{L}(\lambda)$ is unique and is obtained by:*

$$\pi_v = \begin{cases} \infty & \text{if } \lambda = 0 \\ \left(\frac{d}{d\tau}\mathcal{B}_v\right)^{-1}(\lambda r_v) & \text{if } 0 > \lambda \geq -\frac{N_v K_v}{r_v \kappa}\log\left(A_v^* + 1\right) \\ 0 & \text{otherwise} \end{cases}$$

6.6 Solution Methods

We propose three methods to solve problem (6.1) - (6.4). The methods are general in the sense that they are capable of dealing with the complete variety of disease burden progression functions discussed in Section 6.4. In addition, we propose several simple planning policies that are worthy of exploration for practical purposes, where a simple policy is currently the norm.

6.6.1 BLP Approach

The binary linear programming (BLP) approach takes formulation (6.1) - (6.4) as a starting point and tackles the non-linearity of burden function $\mathcal{B}_v(\tau_v)$ by discretizing the disease burden progression function $f_v(s)$. We define a discretization as an ordered set of burden levels $\mathcal{F}_v = \{f_v(s_1), f_v(s_2), \ldots, f_v\left(s_{|\mathcal{F}_v|}\right)\}$, where

$f_v(s_1) < f_v(s_2) < \ldots$. Here, $f_v(s_1)$ is a lower bound on attainable disease burden values.

Let $i \in \mathcal{F}_v$ be the burden level of village v at the beginning of planning period t, and $j \in \mathcal{F}_v$ be the burden level of village v at the end of period t. Notice that j depends on whether village v is screened in period t. Binary parameter A^1_{vij} equals 1 if screening village v, which is at burden level i at the current period's beginning, results in burden level j at the next period's beginning, and 0 otherwise. Similarly, binary parameters A^0_{vij} reflect the burden level transitions in case village v is not screened in period t.

Since we assume each screening round takes place at the end of a planning period (see Section 6.3), the average expected burden incurred *during* period t only depends on the disease burden level i at the beginning of that period. Let parameters $\bar{b}_{vi}$ represent the corresponding average expected burden. Furthermore, let variables z_{vit} indicate whether or not village v encounters burden level $i \in \mathcal{F}_v$ at the beginning of period t. For the beginning of period 1, the expected disease burden level is indicated by binary parameters ζ_{vi1}, and we set z_{vi1} correspondingly. Using this notation, the planning and allocation problem can be formulated as the following BLP problem:

$$\min \sum_{v\in\mathcal{V}} \sum_{i\in\mathcal{F}_v} \sum_{t\in\mathcal{T}} \frac{1}{T} \bar{b}_{vi} z_{vit} \tag{6.24}$$

$$\text{s.t. } z_{vjt+1} \geq z_{vit} + x_{vt} - 1 \qquad \forall (i,j) : A^1_{vij} = 1, v \in \mathcal{V}, \; t \in \{1,\ldots,T-1\} \tag{6.25}$$

$$z_{vjt+1} \geq z_{vit} + (1 - x_{vt}) - 1 \qquad \forall (i,j) : A^0_{vij} = 1, v \in \mathcal{V}, \; t \in \{1,\ldots,T-1\} \tag{6.26}$$

$$\sum_{i\in\mathcal{F}_v} z_{vit} = 1 \qquad \forall v \in \mathcal{V}, t \in \mathcal{T} \tag{6.27}$$

$$z_{vi1} = \zeta_{vi1} \qquad \forall v \in \mathcal{V}, t \in \mathcal{T} \tag{6.28}$$

$$\sum_{v\in\mathcal{V}_c} r_v x_{vt} \leq y_{ct} \qquad t \in \mathcal{T}, c \in \mathcal{C} \tag{6.29}$$

$$\sum_{c\in\mathcal{C}} y_{ct} \leq M \qquad t \in \mathcal{T} \tag{6.30}$$

$$x_{vt}, y_{ct}, z_{vit} \in \{0,1\} \qquad v \in \mathcal{V}, c \in \mathcal{C}, t \in \mathcal{T}, i \in \mathcal{F}_v \tag{6.31}$$

Here, (6.24) - (6.28) model the discretized objective function. The other constraints have the same interpretation as in formulation (6.1) - (6.4). We observe that both the number of constraints and the number of variables are $\mathcal{O}\left(|\mathcal{V}| \cdot |\mathcal{T}| \cdot \max_v\{|\mathcal{F}_v|\}\right)$.

The proposed discretization may restrict the optimization to an incomplete set of relevant disease burden values and hence may imply incorrect solution values and suboptimal solutions. Discretized formulations can, however, be ensured to be exact in the following two ways. First, we can pre-calculate all attainable burden levels and include them into $\mathcal{F}_v$. Since, in principle, the number of possible burden levels grows exponentially with T, this is typically only feasible when T is small. As an alternative, we can repeatedly solve the model while adding the *actual set of burden levels* $\mathcal{F}_{va}$ encountered by each village to $\mathcal{F}_v$. Details about this algorithm are provided in Appendix 6.C. Computation times are also a major drawback of this approach.

6.6.2 Iterated Local Optimization

For a given time period t and solution $(\mathbf{x}, \mathbf{y})$, which satisfies $y_{ct} = 0$ for all $c \in \mathcal{C}$, let $\hat{b}_{ct}(\mathbf{x})$ denote the solution value improvement when choosing to additionally send a mobile team to cluster c in period t. When considering the planning for all other periods to be fixed, it is optimal to send mobile teams to the M clusters for which $\hat{b}_{ct}(\mathbf{x})$ is largest. This idea is elaborated in the iterated local optimization (ILO) approach, as formally explained in Algorithm 2. The algorithm repeatedly selects a period t, sets $y_{ct} = 0$ for all clusters c, and reoptimizes the planning for that period while keeping the planning for all other periods fixed. The solution thus found can be identical to the initial solution, or can be another solution with the same or lower solution value. To improve the quality of the final solutions obtained, we run the algorithm for two starting solutions: $\mathbf{x} = \mathbf{0}$ and the initially infeasible solution $\mathbf{x} = \mathbf{1}$, and continue until no further improvements are found.

Algorithm 2 *Iterated local optimization method.*

```
while the solution changes do
    for t ∈ 𝒯 do
        Fix x_vs for all s ≠ t
        Determine b̂_ct(x) and corresponding decisions x_vt
        Set y_ct = 1 for the M clusters with largest b̂_ct(x) values
        Update x_vt
    end for
end while
```

Determining $\hat{b}_{ct}(\mathbf{x})$ is an optimization problem in itself. For a given solution $\mathbf{x}$ and period t, let $\hat{b}_{vt}(\mathbf{x})$ denote the decrease in average burden when choosing to screen village $v \in V_c$ instead of choosing not to. Then the problem is to find the subset of villages in cluster c that maximizes the total burden decrease while not exceeding screening capacity. We note that this is equivalent to a knapsack problem with values $\hat{b}_v(\mathbf{x})$, weights r_v, and capacity 1. We solve this problem as a BLP (see Kellerer et al., 2004).

6.6.3 Column Generation Approach

An alternative approach is to formulate the problem in terms of selecting a visit schedule or *visit pattern* for each of the villages. Let $\mathcal{P}$ denote the set of patterns, each of which can be characterized by a binary vector $\mathbb{B}^{|\mathcal{T}|}$. Subset $\mathcal{P}_t$ represents the patterns in which a screening round takes place at time t. Furthermore, let variables $x_{\hat{p}v}$ indicate whether pattern $\hat{p}$ is assigned to village v and let $\bar{b}_{\hat{p}v}$ denote the corresponding average expected disease burden. Then the planning and allocation problem can be formulated as:

$$\min Z = \sum_{v \in \mathcal{V}} \sum_{\hat{p} \in \mathcal{P}} \bar{b}_{\hat{p}v} x_{\hat{p}v} \tag{6.32}$$

$$\text{s.t.} \sum_{\hat{p} \in \mathcal{P}} x_{\hat{p}v} = 1 \qquad v \in \mathcal{V} \tag{6.33}$$

$$\sum_{v \in c} \sum_{\hat{p} \in \mathcal{P}_t} r_v x_{\hat{p}v} \leq y_{ct} \qquad c \in \mathcal{C}, t \in \mathcal{T} \tag{6.34}$$

$$\sum_{c \in \mathcal{C}} y_{ct} = M \qquad t \in \mathcal{T} \tag{6.35}$$

$$x_{\hat{p}v}, y_{ct} \in \{0,1\} \qquad \hat{p} \in \mathcal{P}, v \in \mathcal{V}, c \in \mathcal{C} \tag{6.36}$$

A main disadvantage of this formulation is that the number of patterns equals $2^{|\mathcal{T}|}$, and hence that the number of variables grows exponentially with $|\mathcal{T}|$. Our third proposed solution approach is therefore based on column generation.

As is standard practice, our column generation approach starts from solving the so-called master problem (MP) – the LP relaxation of problem (6.32) - (6.36) – using only a subset of the visit patterns. The resulting problem is referred to as the restricted master problem (RMP). Next, "promising" patterns are identified in the so-called pricing problem, and added to the RMP. This process is repeated until no more promising patterns can be found. The resulting set of patterns is then added to (6.32) - (6.36), which can subsequently be solved to optimality using integer programming techniques. Notice that this approach is not necessarily exact, as the optimal solution to (6.32) - (6.36) may require patterns which have not been included or generated to find the optimal solution to the LP relaxation.

Pricing Problem. For a given solution, the pricing problem for village v corresponds to finding a pattern with negative reduced costs. Let c_v denote the cluster village v belongs to and let γ_v and γ_{ct} represent the dual variables corresponding to constraints (6.33) and (6.34). Then the reduced costs of column $\hat{p}$ are given by (cf. Desaulniers et al., 2005):

$$rc_{\hat{p}v} = \bar{b}_{\hat{p}v} - \gamma_v - \sum_{t:\hat{p}\in\mathcal{P}_t} r_v \gamma_{c_v t}. \tag{6.37}$$

Consequently, the pricing problem is defined as:

$$rc_v^* = \min_{\hat{p}\in\mathcal{P}} \bar{b}_{\hat{p}v} - \gamma_v - \sum_{t:\hat{p}\in\mathcal{P}_t} r_v \gamma_{c_v t}. \tag{6.38}$$

LP duality implies the following lower bound on the solution value of the LP relaxation of our problem, Z_{MP}, and hence of Z (Desaulniers et al., 2005). Here, Z_{RMP} denotes the solution value of the RMP:

$$Z_{RMP} + \sum_{v\in\mathcal{V}} rc_v^* \leq Z_{MP} \leq Z \tag{6.39}$$

In solving the pricing problem, we again encounter the difficulties caused by the non-linearity of the burden progression function. We deal with this as follows. First, we again discretize the progression function $f_v(s)$, yielding the set of burden levels $\mathcal{F}_v$. Next, we build the graph $\mathcal{G}(\mathcal{F}_v)$ depicted in Figure 6.4. Each column of nodes represents one period, and each node within a column represents a burden level $i \in \mathcal{F}_v$. A visit pattern corresponds to a sequence of the red and blue arcs which model the sequence of screening decisions. A blue (red) arc from node i in period t to node j in period $t+1$ represents the situation where village v has burden level i at the beginning of period t and is screened (not screened) in period t, resulting in burden level j at the beginning of period $t+1$. As the node set forms a discretization of the possible burden levels, the burden level of node j may not exactly equal the burden level j^* resulting from screening at burden level i. We choose $j = \max_{j'\in\mathcal{F}_v : j'\leq j^*} j'$, i.e. the node with highest burden level not larger than j^*. Hence the resulting burden level j is a lower bound for the actual resulting burden level. Parameter $\bar{b}_i$ denotes

the average expected disease burden in a given period when burden level i represents the disease burden at the beginning of the period.

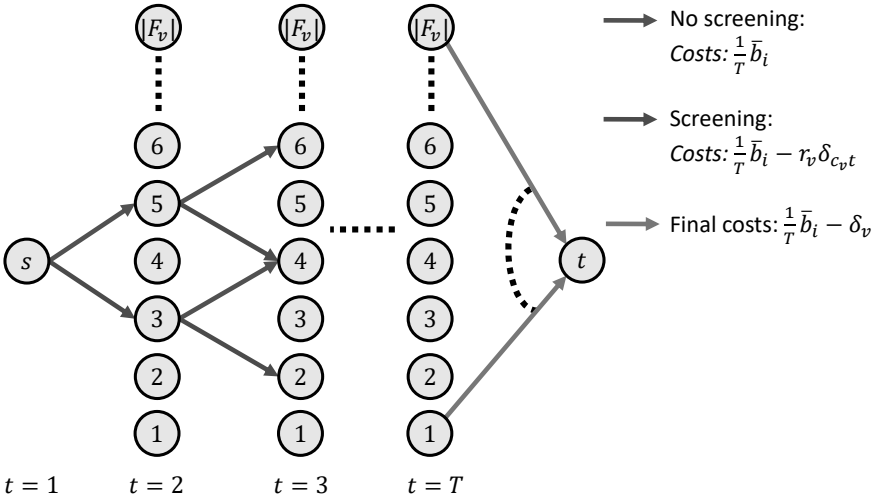


Figure 6.4: Graph used to solve the pricing problem for village v as a shortest path problem.

It can easily be verified that the length of a given $s-t$ path provides a lower bound on the reduced costs of the corresponding visit pattern. Consequently, the shortest path obtained for any discretization $\mathcal{F}_v$ yields a lower bound on rc_v^* and hence can serve to provide a lower bound on Z (see equation (6.39)). Moreover, if the burden levels visited by a given $s-t$ path mirror the actual burden levels encountered, its length exactly equals reduced costs. As we motivate in Appendix 6.D, this implies that one can obtain an exact solution method for the pricing problem by repeatedly solving the shortest path problem.

Heuristic Pricing. A common strategy to accelerate a column generation approach is to search for *one negative* reduced costs column rather than *the most negative* one (Desaulniers et al., 2002). Since many iterations tend to be needed to find the optimal pattern using the algorithm presented in Appendix 6.D, we also follow this strategy. Specifically, we first try to find a promising pattern using the local

search algorithm explained in Appendix 6.E. If this fails, we apply the shortest path approach for a dense discretization (see Figure 6.4).

Warm Start. A second acceleration strategy is to give the column generation algorithm a "warm start" by providing it with the patterns from the solutions obtained by iterated local optimization.

6.6.4 Planning Policies

We now propose four simple planning policies which generate solutions period by period, starting from period one, using straightforward polynomial time heuristics.

1. The *equalization* policy equalizes screening frequencies of the villages. We define $\tau_u = \sum_v r_v/M$, i.e. the minimum number of planning periods needed to visit all villages once. Then, in period t, we first determine for each cluster c the number $N(t, c)$ of people who were not screened in the past τ_u planning periods. Next, the policy selects for screening in period t the M clusters with highest $N(t, c)$. Within each cluster, the teams are assigned to villages in descending order of the time since the last screening round, until capacity is consumed.

2. The *filtering* policy applies the equalization policy but restricts the problem instance to a subset of the villages for which screening is considered most urgent.

3. The *differentiation* policy generalizes the filtering policy and the equalization policy. It first classifies villages into several classes ec (e.g., high and low), depending on an urgency measure. In addition, it assigns to each class ec a target screening interval τ_{ec}. Next, in period t, we firstly determine for each cluster c the number $N(t, c)$ of people for which the time since the last screening equals at least their target interval. Assignment of teams to clusters and to villages within the clusters is done as in the equalization policy.

4. The *burden level* policy strives to screen the villages with the highest burden levels. For each cluster c, it calculates $l(c)$, the total expected disease burden at the beginning of the period for the persons screened in that cluster if a team

is assigned to it. The specific persons screened are determined by selecting villages in decreasing order of expected disease burden per person until capacity is consumed. Next, the policy assigns teams to the M clusters with highest $l(c)$.

6.7 Case Study

This section numerically illustrates the application of our results and methods to HAT, the disease which motivated our work. The presented case study is based on a database of screening operations and villages in Kwamouth, DRC. Using available data and models (Chapter 5), we analyze the computational and solution quality for the presented solution approaches, as well as their sensitivity with respect to data impreciseness. We implemented our experiments in Matlab R2015a and used CPLEX 12.63 as a BLP solver.

6.7.1 Baseline Case Description

Our dataset describes HAT screening rounds in 2324 villages in the Kwamouth health district (Bandundu, DRC) between 2004 and 2013. From these villages, 239 were included in our case study based on two criteria: (1) there exists at least one record of the number of people screened and (2) the geo-coordinates are known. The first criterion is required to estimate population sizes and the second is required for assigning villages to clusters. Participation (*part*) in screening rounds in Bandundu has been estimated as 71% (Robays et al., 2004). If this percentage were always attained exactly, the total population equals 1.41 times the number of people participating. However, participation varies substantially (Robays et al., 2004). Thus, we estimate the total village population to be 1.2 times the maximum number of people participating in a screening round reported for that village. To ensure that each village can be screened in one planning period, one village had to be split into two. Both are considered separate villages in the remainder of our analyses, which increases the total number of villages to 240.

Current planning practices in the national sleeping sickness control program of the DRC largely support cluster-based planning (Hasker, 2016):

> *"They define different axes, and then simply visit one axis per trip. (...) An axis can be a major road or a river by which you travel. Of course, there are not many roads. Most villages can only be accessed by one road, so if you are travelling in a given direction, it is logical to stay in that region."*

> *"The planning assumes that they screen 300 persons per day, 20 days per month, so 6000 [persons] per month, and then they return to their basis."*

Reflecting current practice, we let a planning period represent one month. We estimate that the number of people participating in a screening round in village v equals $part \cdot N_v$ and that each screening round lasts a fraction $r_v = part \cdot N_v/6\,000$ of a planning period. We initially consider the case of one screening team. Since the total estimated number of people from the 240 villages participating in screening equals 73 521, this team would need approximately 12 months to visit the villages, which is in line with the current WHO guidelines for villages having one case in the past three years (WHO, 2015b). We manually clustered the villages, following the structure of the road network. The resulting clusters, the villages, and their relative population sizes are depicted in Figure 6.5.

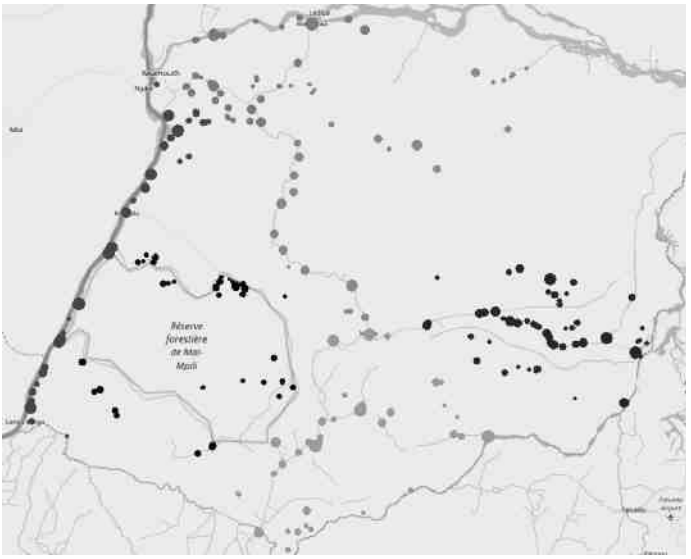


Figure 6.5: Map of the 240 villages in Kwamouth included in our case study. Population is indicated by node size.

Average expected number of people infected, as calculated by (6.10), is used as an indicator of disease burden. Carrying capacities for 70 villages were obtained from De Vries et al. (2016). Since sufficient screening data were lacking for the other villages – defined as having information about at least two screening rounds *and* at least one disease case – these were randomly generated from an exponential distribution with mean 0.858%. This distribution was based on 143 carrying capacities estimated for this region by De Vries et al. (2016). We do not claim these estimates to be accurate, but consider them to be realistic enough for the purpose of the computational analysis presented in this case study.

In line with common practice in the evaluation of active case finding (Robays et al., 2004; Piot, 1967), we calculate the impact fraction p as the product of the following fractions: (1) participation, $part$, (2) sensitivity of the screening test, $sens_{scr}$, (3) sensitivity of the confirmation test, $sens_{con}$, and (4) treatment initiation, $treat$. Here, sensitivity is defined as the fraction of infected people who are detected when

participating in screening. Baseline values for these and other parameters used in the case study can be found in Table 6.1. We note that time is measured in months.

Parameter	Value	Source/remark
$part$	0.71	(Results on Bandundu by Robays et al., 2004)
$sens_{scr}$	0.95	(Robays et al., 2004)
$sens_{con}$	0.75	(Lutumba et al., 2006)
$treat$	0.99	(Results on Bandundu by Robays et al., 2004)
κ	0.0667	(De Vries et al., 2016). Adapted from year-based estimate.
A_{v0}	2.526	(De Vries et al., 2016)
M	1	
T	36	

Table 6.1: Baseline case parameters.

6.7.2 Computational Experiments

We now consider the computation times and solution quality for the three solution methods proposed in Section 6.6. To allow the calculation of the exact solution and hence of exact optimality gaps, we first consider small instances in which T ranges from four to seven months (for larger horizons, finding the guaranteed optimal solution becomes extremely difficult) and M ranges from one to two. The BLP approach uses a discretization consisting of 25 prevalence levels equally spaced between zero and the village's carrying capacity. The exact solution was determined by applying this approach for a discretization containing all 2^T prevalence levels that could possibly be attained. Table 6.2 shows the CPU times (sec.) and optimality gaps for the three methods.

		$M = 1$				$M = 2$			
	T	4	5	6	7	4	5	6	7
CPU time (sec.)	BLP	3.0	18.2	62.1	193.0	3.9	30.9	85.5	371.8
	ILO	0.4	0.5	0.7	0.7	0.4	0.8	1.4	2.4
	CG	3.3	4.6	6.4	9.3	3.1	5.0	7.1	9.7
Opt. gap	BLP	0.000%	0.000%	0.000%	0.008%	0.004%	0.005%	0.022%	0.037%
	ILO	0.000%	0.000%	0.000%	0.001%	0.003%	0.006%	0.009%	0.007%
	CG	0.000%	0.000%	0.000%	0.001%	0.003%	0.006%	0.009%	0.007%

Table 6.2: Computational results.

Each of the methods yields optimal or near-optimal solutions. Yet, computation times for the BLP approach soon become impractical, even though the discretization contains only 25 prevalence levels. For the other approaches, solution times grow much more slowly, which renders them more suitable for large problem instances, such as instances considering up to 36 months, as in the case study at hand.

To assess the added value of the column generation approach over the ILO approach, we also applied the three methods to $T \in \{12, 18, 24, 30, 36\}$. In each case, the approaches attain the same solution value, showing that it is hard for the column generation (CG) approach to improve upon ILO solutions. Implications for the potential role of the column generation approach are discussed in the final section. Due to its computational performance and solution quality, we shall use the ILO approach as a solution method in the remainder of this section. Since we have no proof of optimality for the obtained solutions, we refer to them as "optimized" rather than "optimal".

6.7.3 Baseline Case Results

To gain intuition regarding the optimized planning decisions, we now analyze them in detail for the baseline case ($M = 1$, $T = 36$). Figure 6.6 shows how numbers of screening rounds relate to carrying capacities (i.e., node sizes). We observe that there is substantial correlation between the two, as also shown on a more aggregate level in Figure 6.7. Characteristics of other villages within a cluster form a second major determinant. For example, the total population of the blue cluster is comparatively small (see Figure 6.5). As a consequence, it is relatively unattractive to visit this cluster often, which constrains the number of visits to villages in that cluster having a high carrying capacity. The reverse is true as well: several villages with a relatively low carrying capacity are frequently visited due to high carrying capacities of other villages in the same cluster. Next to the need for choosing the clusters deliberately, this also touches upon equity considerations.

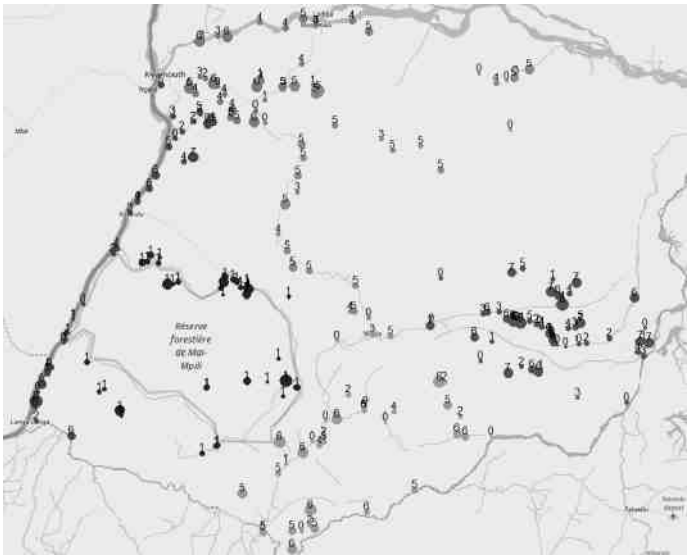


Figure 6.6: Number of screening visits to each village in 36 months. Carrying capacities are indicated by node sizes.

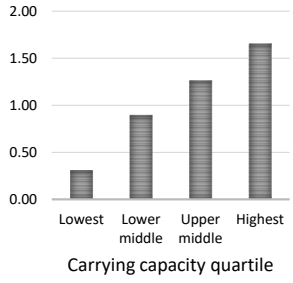


Figure 6.7: Average yearly screening frequencies in the optimized solution for four classes of villages.

A third observation is that 26 out of the 240 villages are not visited in the 36 months considered. In terms of current planning guidelines, this is equivalent to being excluded from the screening program (WHO, 2015b). While exclusion may be preferable from a medium term effectiveness viewpoint, it raises equity concerns, and

can also be problematic when the objective is disease elimination. We come back to these issues in the discussion.

6.7.4 Performance of Planning Policies

Let us now investigate the trade-off between solution quality and solution method complexity. Figure 6.8 describes the performance of the planning policies proposed above in terms of average expected number of people infected. Here, the filtering policy includes all villages assigned a strictly positive frequency in Proposition 6.4 in the screening program. The differentiation policy divides villages into 4 equally sized classes. The first class contains the villages with highest carrying capacities, etc. The policy assigns the classes 40%, 30%, 20%, and 10% of the screening capacity, respectively.

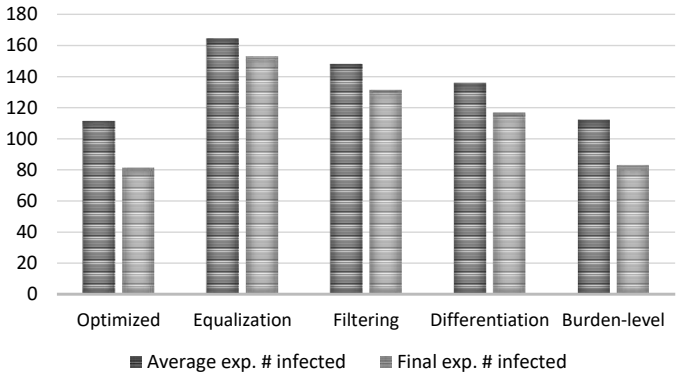


Figure 6.8: Average and final (end of planning horizon of 36 months) expected number of people infected in the 240 villages for the optimized schedule and the schedules following from the planning policies.

Without screening, the average expected number of people infected in the 240 villages during the next 36 months would be 341 persons. Hence, the solutions avert 52% (equalization) up to 67% (optimized) of average prevalence, which shows the substantial impact active case finding can have. We observe that the burden level policy performs only 0.7% worse than optimized planning. The filtering policy

and the differentiation policy perform substantially worse. This is explained by the fact that the optimal screening frequency for a given village is, as discussed, largely determined by the characteristics of villages sharing the same cluster, which these policies do not appropriately account for.

The results more generally show that differentiating screening frequencies rather than uniformly allocating capacity pays off substantially. For example, the average expected number of people infected decreases from 165 persons for the schedule following from the equalization policy to 111 persons (-32%) for the optimized schedule.

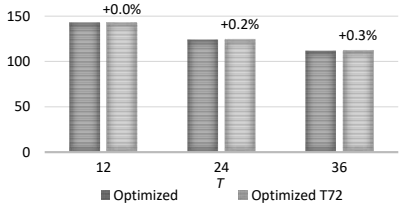
One possible concern about minimizing average disease burden for a short time horizon is that the resulting screening policy may yield an undesirable final disease burden. This potential drawback does not occur in the presented case study. Instead, as shown in Figure 6.8, the average expected number of people infected is well correlated with the final expected number of people infected. Next, we analyze end-of-horizon effects in detail.

6.7.5 End of Horizon Effects

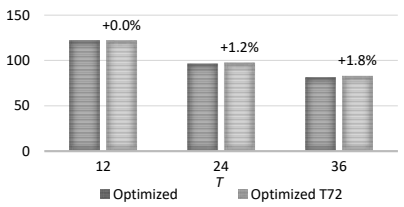
To assess the extent to which planning decisions are determined by the planning horizon, we obtain the optimized solution for $T = 72$ and compare the first twelve months of this solution with the optimized solution for $T = 12$. This is repeated for $T = 24$ and $T = 36$. Figure 6.3 depicts the results. We observe little difference between average disease burdens. Compared to the partial solutions, referred to as "optimized $T72$", average disease burden is slightly lower. Solutions are also comparable in terms of the final expected number of people infected. This indicates that solution value and final disease burden are relatively insensitive to the length of the planning horizon, at least for this example.

6.7.6 Sensitivity Analysis: Carrying Capacities

As we cannot directly observe the carrying capacities of the villages, we estimate them from observables. Chapter 5 proposes and fits a formula relating the carrying capacity to the average observed prevalence and the average screening frequency in the past five years. Due to stochasticity in prevalence levels, however, such estimates



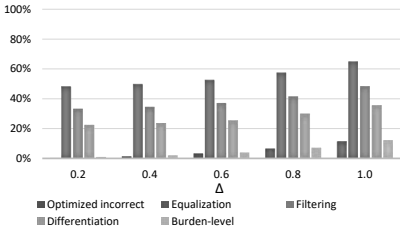
(a) Average expected number of people infected.



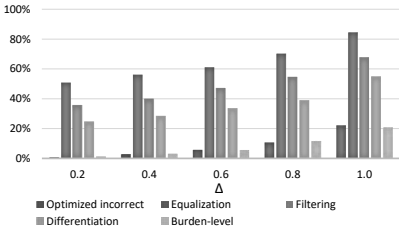
(b) Final expected number of people infected.

Figure 6.9: Comparison of solution for $T \in \{12, 24, 36\}$ with the partial solution for $T = 72$.

may be imprecise, which begs the question to what extent this impacts the quality of scheduling decisions. Figure 6.10 summarizes sensitivity analysis results. For each village, we randomly draw the "real" carrying capacities according to a uniform distribution on $[K_v(1-\Delta), K_v(1+\Delta)]$, determine the "real" optimized solution, and calculate the "real" value of the solutions that were based on the "incorrect" carrying capacities. This is repeated 100 times for each $\Delta \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$ which yields the depicted average and maximum observed optimality gaps.



(a) Average optimality gap (%) based on 100 draws.



(b) Maximum optimality gap (%) in 100 draws.

Figure 6.10: Results of the sensitivity analysis on carrying capacity estimates.

The results show that solution quality is rather robust with respect to impreciseness. For example, the estimated "optimality" gap for the burden level policy ranges from 1.0% for $\Delta = 0.2$ to 12.2% for $\Delta = 1.0$. Hence, even when the real carrying capacities deviate up to 100% from the assumed values, the estimated average gap for

this policy equals only 12.2%. Furthermore, the maximum observed "optimality" gap is rather close to the average. For the solution obtained by the burden level policy, for example, the maximum ranges from 1.5% for $\Delta = 0.2$ to 20.3% for $\Delta = 1.0$. Note that the actual optimality gap may be larger since we used the optimized solution for comparison. We note further that one needs to be careful in generalizing these results. They are based on a specific assumption on the distribution of carrying capacities, and will be affected when this distribution differs in reality. Directions for future research are addressed in the final section.

6.7.7 Sensitivity Analysis: Screening Impact

There is an ongoing debate about the expected impact of active case finding (Welburn et al., 2016), which is known to vary among regions (Robays et al., 2004) and may change over time. As a consequence, the true impact of screening may deviate from the assumed impact, as quantified by parameter p in the presented fractional impact models. Figure 6.11 depicts how the quality of solutions obtained for baseline value $p = 0.5$ (see Table 6.1) is affected when p equals 0.1, 0.3, 0.7, or 0.9 in reality. Here, "optimized incorrect" refers to the optimized solution using $p = 0.5$. To estimate the resulting optimality gap, we also present the solution value for the optimized solution using the "real" value of p.

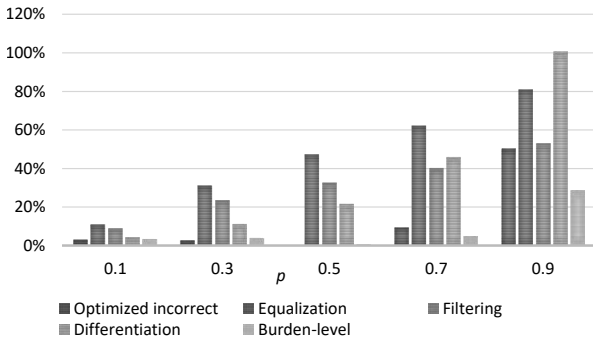


Figure 6.11: Optimality gap attained when basing decisions on the baseline value for p, for different actual values of p.

We observe that the estimated optimality gap for the optimized schedule using $p = 0.5$ is only 3.1%, 2.7%, and 9.4% when in reality p equals 0.1, 0.3, and 0.7, respectively. This shows that even a substantial over- or underestimation of impact does not necessarily have serious consequences. When p equals 0.9 in reality, however, sub-optimality increases to 50.2%.

A second observation is that the burden level policy outperforms optimized planning by 4.5 and 21.4 percent points when p equals 0.7 and 0.9, respectively. This strengthens the belief that the burden level policy provides a good alternative to optimized planning. The schedules obtained by the equalization, filtering, and differentiation policies remain inferior to the optimized schedules, irrespective of p.

6.8 Conclusions and Discussion

This chapter considers the problem of scheduling mobile population screening teams. Disease burden in a given village or population is assumed to increase between screening rounds. Screening is modeled to decrease the disease burden either by a given fraction or to some fixed level. We present generic solution methods and insights, which have much potential to be of value in solving practical problems. This is illustrated through the case study of HAT disease control in Kwamouth, DRC.

One main takeaway is that effectiveness of screening programs can increase substantially when rationalizing planning decisions based on endemicity and clustering. For example, we estimate that the average expected number of people infected in the region considered decreases from 165 persons to 111 persons (-32%) when using an approach that incorporates these aspects instead of an approach that simply assigns (approximately) the same screening frequency to each village. This has important implications for current guidelines. The WHO recommends yearly screening for villages which had at least one case in the past three years screening every three year for villages which had no case in the last three years but at least one case in the past five years (WHO, 2015b). However, screening capacity may not be sufficient to implement these guidelines. Moreover, our results suggest that the lack of differentiation can result in substantial sub-optimality and hence the WHO guidelines could likely be improved by aligning screening frequencies more closely with prevalence.

A second major finding is that simple planning policies yield solutions that are close to those obtained by more sophisticated methods. For example, a policy that solely bases decisions on current disease burden levels performs only 0.7% worse than methods employing sophisticated optimization techniques. This is important, since complexity tends to go hand in hand with costs in terms of software implementation, software maintenance, and/or training (cf. Banker et al., 1993). From a cost-effectiveness point of view, choosing such a simple policy might thus be preferable. One should note, though, that the most suitable method may differ by context. Our numerical results were obtained for a specific disease burden progression function and a specific set of villages and might change for other case studies. Moreover, specific applications might require adaptations from the presented solution methods, and some might be more suitable for such adaptations than others. For example, if policy makers impose a maximum screening interval per village, this could easily be incorporated in pricing problem (6.39).

Third, sensitivity analyses for our case study suggest that solution quality is relatively insensitive to impreciseness of input data, at least for the sophisticated methods and the burden level policy. Due to inherent stochasticity of infectious disease outbreaks and screening impact, this is very important. When estimates become very imprecise, however, substantial sub-optimality may be induced, which shows the necessity of investing in reliable parameter estimates. Future research testing our policies and solution approaches within a validated simulation model would be highly valuable. Furthermore, one could investigate models that explicitly account for impreciseness. Bayesian statistics, for example, could be very well used to dynamically update beliefs about the burden progression function, given newly obtained burden level estimates. Incorporating such beliefs would make the problem stochastic, and may require different models and solution methods.

To the best of our knowledge, we are the first to present generic methods to analyze the relationship between capacity and long-term average disease burden. These can assist policy makers, e.g., in determining how many teams are expected to be needed to stabilize the epidemic or to achieve disease eradication. For some cases, the analyses reveal analytical expressions for this relationship. For example, we prove that the function is piecewise linear for the logistic model with fractional screening

impact. We note, though, that the analyses are based on a problem variant in which clustering constraints are relaxed, which suggests that the estimated capacity requirements for reaching a given disease burden target level are lower bounds on the actual capacity requirements. This is not completely trivial, since we restrict ourselves to fixed time intervals between screening rounds. Though this has been proven optimal for a special case of $MSTD_R$ (Bar-Noy et al., 2002), we have not been able to do so for the general setting considered. Future research is needed to investigate this.

A limitation of our methods is that they assume villages to be independent: intervening in one village or deciding not to do so does not impact the epidemic in others. Since infectious disease agents can move between different villages, this potentially affects solution quality. In our case study, we do not expect this to be substantial, since tsetse flies do not cover large distances (Williams et al., 1992) and since poor infrastructures in the DRC severely restrict human mobility (Foster and Benitez, 2011). Nevertheless, in a situation in which elimination is close, as is the case in many HAT disease foci (Simarro et al., 2011), one may wish to also visit villages that are of low endemicity, so as to reduce the risk that the disease re-emerges from them. Equity considerations might form a second reason to provide some level of access to screening in such villages (cf. Ares et al., 2016; McCoy and Lee, 2014). Future research is needed to propose models and solution methods incorporating such aspects.

Appendices

6.A Proofs

Proof of Proposition 6.1.

Proof. This directly follows from the fact that $MSTD$ has the PMP (see Section 6.2) as a special case, which was shown to be NP-hard by Grigoriev et al. (2006). □

Proof of Proposition 6.2.

Proof. The integral in (6.8) is the convolution of the functions $\lambda_v(s)$ and $\beta(s)$, denoted by $\lambda_v * \beta$. Taking the derivative of a convolution gives (Bracewell, 1965):

$$(\lambda_v * \beta)' = \lambda_v' * \beta = \lambda_v * \beta'$$

Hence, this yields:

$$\frac{d}{d\tau} f_v(\tau) = \int_0^\tau \lambda_v'(s)\beta(\tau - s)ds \tag{6.40}$$

$$= \int_0^\tau \lambda_v(s)\beta'(\tau - s)ds. \tag{6.41}$$

We know that $\lambda_v \geq 0$ and $\beta \geq 0$. Moreover, it is given that $\lambda_v' > 0$ and/or $\beta' > 0$. This implies that the derivative of function (6.8) is strictly positive and hence this function is increasing.

□

Proof of Proposition 6.3.

Proof. Consider a 3-partition instance with target value B and positive integers $B/4 < r_i < B/2$ for $i \in \{1, ..., 3T\}$ and a target value B such that $\sum_i r_i = TB$. The 3-partition problem requires one to decide whether the integers can be partitioned into T triples (i, j, k) such that $r_i + r_j + r_k = B$ for each of the T triplets. The 3-partition problem is known to be strongly NP-complete (Gary and Johnson, 1979).

A polynomial-time reduction from 3-partition to $MSTD_{HAT}$ is obtained as follows. We consider $T + 1$ planning periods, one cluster of villages, and one screening team. For each integer $i \in \{1, ..., 3T\}$, we introduce a village $v(i)$, resulting in $3T$ villages. For $i \in \{1, ..., 3T\}$, village $v(i)$ requires screening capacity $r_{v(i)} = r_i/B$. Moreover, village $v(i)$ has population $N_{v(i)} = r_i$, carrying capacity $K_{v(i)} = 1$, and $A_{v(i)0} = 0$. There is one team available. The reduction will simultaneously cover the

cases of fixed screening impact and fractional screening impact. For the fractional impact, we choose $p = 1$. For the fixed impact, we equivalently set $A_v^* = 0$. These parameter settings imply that prevalence in a given village $v(i)$ equals 1 until the first period in which $v(i)$ is screened and is reset to zero starting from the end of that first screening period (because of assumption 2).

We now claim that the 3-partition instance I is a *yes*-instance if and only if the $MSTD_{HAT}$ instance has a solution of value at most $\frac{1}{2}T(T+1)B$. The proof is straightforward and the main intuition is depicted in Figure 6.12.

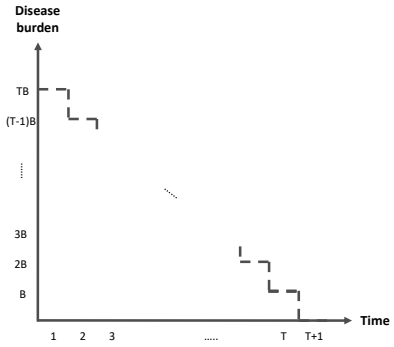


Figure 6.12: Disease burden incurred over time in the $MSTD_{HAT}$ instance if the 3-partition instance is a *yes*-instance.

If. Let $S = (s_1, s_2, \ldots, s_T)$ be solution for I satisfying $r_i + r_j + r_k = B$ for each triplet $s_t, t = 1, \ldots, T$. Then a solution for $MSTD_{HAT}$ is formed by simply maintaining the triplets of corresponding villages $(v(i), v(j), v(k))$. Now, schedule villages corresponding to the first triplet s_1 at period 1, the villages corresponding to the second triplet at period 2, et cetera, until the villages of the final triplet s_T are schedule at period T. Notice that the screening of the villages corresponding to integers in triplet s_t is assumed to take effect at time $t, t = 1, \ldots, T$.

The initial disease burden at the beginning of the planning horizon is given by $\sum_{i=1}^{n} N_{v(i)}$, which in turn equals TB. Hence this is also the total burden during this period, as screenings takes effect when the period ends. More generally, during period t, the scheduling of villages $v(i) \in s_t$ reduces the burden by $\sum_{i \in s_t} N_{v(i)} = B$ a

time t, while the burden of disease during period t sums to $(T-t+1)B$, as depicted in Figure 6.12. Hence the total disease burden from time 0 to time $T+1$ equals $\sum_{t=1}^{T+1}(T-t+1)B = \sum_{t=1}^{T} tB = \frac{1}{2}T(T+1)B$.

Only if. Now suppose the $MSTD_{HAT}$ instance has a solution of value at most $\frac{1}{2}T(T+1)B$. As there is one team available in each period $t, t = 1, \ldots, T$, the sum of the screening capacity consumptions $r_{v(i)} = r_i/B$ of the villages $v(i)$ screened in period t cannot exceed 1. As a result, a population of at most B people can be screened in each period. Starting with the given disease burden of TB, and reducing the disease burden per period with the maximum attainable value of B, we arrive at a total minimum disease burden of $\sum_{t=1}^{T+1}(T-t+1)B = \sum_{t=1}^{T} tB = \frac{1}{2}T(T+1)B$, as is required for a *yes*-instance of $MSTD_{HAT}$. This value is only attained if in each period $t, t = 1, \ldots, T+1$, the reduction in disease burden is equal to the maximum possible reduction per period of B. As $r_{v(i)} = r_i/B$, it then follows from the fact that $B/4 < r_i < B/2$ for $i \in \{1, ..., 3T\}$, that exactly 3 villages are screened in each period $t, t = 1 \ldots, T$. As the sums of the screening capacity consumptions of the villages $v(i), v(j), v(k)$ screened in the same period t add up to 1, $r_i + r_j + r_k = B$. Hence, the T triplets of the village indices form a (certificate for a) *yes*-answer for 3-partition instance I.

Note that the decision version of $MSTD_{HAT}$ is in NP since we can calculate the value of a solution in at most $|\mathcal{V}| \cdot |\mathcal{T}|$ steps: for each village and for each of the time intervals τ_{vn}, one needs to determine A_{vn}, which costs a constant amount of time. Next, the solution value is calculated by means of (6.10). This completes the proof that the decision version of $MSTD_{HAT}$ is strongly NP-complete.

□

Proof of Lemma 6.1.

Proof. First, note that the fact that $f_v(s)$ is increasing implies that $sgn\left(f_v(s_1) - f_v(s_2)\right) = sgn\left(f_v(s_1+\tau) - f_v(s_2+\tau)\right)$ for any pair of values s_1 and s_2 and any non-negative constant τ. Furthermore, observe that the relationship between $f_v(s_{vn+1})$ and $f_v(s_{vn})$, as described in (6.5), implies that:

$$f_v(s_{vn+2}) - f_v(s_{vn+1}) = (1-p)\left(f_v(s_{vn+1}+\tau_v) - f_v(s_{vn}+\tau_v)\right),$$

Combining these observations yields that $sgn\left(f_v(s_{vn+2}) - f_v(s_{vn+1})\right) = sgn\left(f_v(s_{vn+1}) - f_v(s_{vn})\right)$ Applying this argument repeatedly completes the proof. □

Proof of Lemma 6.2.

Proof. Let functions $g_v(s)$ and $G_v(\tau)$ be defined as follows:

$$g_v(s) = f_v\left(s_v^* + s\right), \tag{6.42}$$

$$G(\tau) = \int_0^{\tau} g_v(s)ds. \tag{6.43}$$

Using this notation, function $\mathcal{B}_v\left(\frac{1}{\pi_v}\right)$ can be written as $\pi_v G_v\left(\frac{1}{\pi_v}\right)$. The fact that this function is decreasing follows from a geometric interpretation of its derivative:

$$\frac{d}{d\pi_v}\pi_v G_v\left(\frac{1}{\pi_v}\right) = G_v\left(\frac{1}{\pi_v}\right) - \frac{1}{\pi_v}g_v\left(\frac{1}{\pi_v}\right)$$

The first term represents the area under the curve $g_v(s)$ for the domain $[0, \frac{1}{\pi_v}]$. The second represents a box of height $g_v(\frac{1}{\pi_v})$ and width $\frac{1}{\pi_v}$. Because $g_v(s)$ is increasing and non-negative (since $f_v(s)$ is), the second part strictly contains the first, which implies that the first derivative is negative on this domain.

Convexity of $\mathcal{B}_v\left(\frac{1}{\pi_v}\right)$ follows from the facts that $\pi_v G_v\left(\frac{1}{\pi_v}\right)$ represents the so-called *perspective* of $G_v(\tau_v)$, that $G_v(\tau_v)$ is convex since $g_v(\tau_v)$ is increasing (Hardy et al., 1952), and that the perspective of a convex function is convex as well for $\pi_v > 0$ (Boyd and Vandenberghe, 2004). □

Proof of Lemma 6.3.

Proof. Our summation is the Cesaro mean of the sequence $\{\bar{f}_{vn}\}_n$. Chapter 5 shows that this sequence monotonically converges to the value defined in (6.17). The fact

that the Cesaro mean of a convergent sequence yields the limit value when $n \to \infty$ (Hardy, 2000) proves our result. □

Proof of Lemma 6.4.

Proof. Suppose there exists an optimal solution with $\tau_v < \tau^*$. Then increasing τ_v to τ^* does not change the value of the burden function (6.17) and does not violate capacity constraint (6.14). The fact that $\tau^* > 0$ follows from the given that $p > 0$. □

Proof of Proposition 6.4.

Proof. Problem (6.18) - (6.20) can be seen as a continuous knapsack problem with capacity M, items v, weights r_v, and values $\frac{-N_v K_v \log(1-p)}{\kappa}$. The optimal solution to this problem is to "select" items in descending order of the ratio of value over weight (Kellerer et al., 2004). This corresponds to ordering the villages in descending order of the presented ratio, and (in this order) setting $\pi_v = \pi^*$ (i.e., $\tau_v = \tau^*$) if remaining capacity suffices and setting π_v to minimum possible screening frequency otherwise. □

Proof of Lemma 6.5.

Proof. For the logistic model, functions $g_v(s)$ and $G(\tau)$ (see (6.42) and (6.43)) are defined as:

$$g_v(s) = \frac{N_v K_v}{1 + A_v^* e^{-\kappa s}} \tag{6.44}$$

$$G(\tau) = \frac{N_v K_v}{\kappa} \log\left(\frac{A_v^* + e^{\kappa\tau}}{A_v^* + 1}\right) \tag{6.45}$$

$$\tag{6.46}$$

Using this notation, function $\mathcal{B}_v\left(\frac{1}{\pi_v}\right)$ can be calculated as $\pi_v G_v\left(\frac{1}{\pi_v}\right)$. Furthermore, the chain rule implies that its first derivative is calculated as:

$$\frac{d}{d\pi_v}\pi_v G_v\left(\frac{1}{\pi_v}\right) = G_v\left(\frac{1}{\pi_v}\right) - \frac{1}{\pi_v} g_v\left(\frac{1}{\pi_v}\right)$$

The two terms have a geometric interpretation. The first represents the area under the curve $g_v(s)$ for the domain $[0, \frac{1}{\pi_v}]$. The second represents a box of height $g_v(\frac{1}{\pi_v})$ and width $\frac{1}{\pi_v}$. Since $g_v(s)$ is strictly increasing (see equation (6.9)), the second part strictly contains the first for all $\pi_v > 0$, and this implies that the first derivative is negative on this domain.

The following derivation shows that the function is also decreasing at $\pi_v = 0$. First, we define the boundary value $\mathcal{B}_v(\frac{1}{0})$ as:

$$\begin{aligned}\mathcal{B}_v\left(\frac{1}{0}\right) &= \lim_{\pi_v \downarrow 0} \frac{N_v K_v \pi_v}{\kappa} \log\left(\frac{A_v^* + e^{\frac{\kappa}{\pi_v}}}{A_v^* + 1}\right) \\ &= \lim_{\pi_v \downarrow 0} \frac{N_v K_v \pi_v}{\kappa} \left(\frac{\kappa}{\pi_v} - \log\left(A_v^* + 1\right)\right) \\ &= N_v K_v\end{aligned}$$

This is then used to calculate the derivative at $\pi_v = 0$:

$$\begin{aligned}\frac{d}{d\pi_v}\mathcal{B}_v\left(\frac{1}{0}\right) &= \lim_{\varepsilon \downarrow 0} \frac{\varepsilon \mathcal{B}_v\left(\frac{1}{\varepsilon}\right) - \mathcal{B}_v\left(\frac{1}{0}\right)}{\varepsilon} \\ &= \lim_{\varepsilon \downarrow 0} \frac{N_v K_v}{\varepsilon} - \frac{N_v K_v}{\kappa} \log\left(A_v^* + 1\right) - \frac{N_v K_v}{\varepsilon} \\ &= -\frac{N_v K_v}{\kappa} \log\left(A_v^* + 1\right) \quad < \quad 0 \end{aligned} \tag{6.47}$$

Differentiating $\mathcal{B}_v\left(\frac{1}{\pi_v}\right)$ twice yields:

$$\frac{d^2}{d\pi_v^2}\mathcal{B}_v\left(\frac{1}{\pi_v}\right) = \frac{A_v^* \kappa N_v K_v e^{\frac{\kappa}{\pi_v}}}{\pi_v^3 \left(A_v^* + e^{\frac{\kappa}{\pi_v}}\right)^2}.$$

Since all terms in the numerator and denominator are positive, this proves the convexity of the function. □

Proof of Proposition 6.5. We first prove the following lemma:

Lemma 6.6. *The range of the function* $\frac{d}{d\pi_v}\mathcal{B}_v\left(\frac{1}{\pi_v}\right)$ *is* $\left[-\frac{N_v K_v}{\kappa}\log\left(A_v^*+1\right),0\right)$

Proof. Note that $\mathcal{B}_v\left(\frac{1}{\pi_v}\right)$ is only defined for $\pi_v \geq 0$. By Lemma 6.5, we know that the derivative of this function is negative and increasing on this domain. Hence, we only need to analyze the extremes of the domain. The lower bound on the range follows from Lemma 6.5, which shows that the derivative at $\pi_v = 0$ equals $-\frac{N_v K_v}{\kappa}\log\left(A_v^*+1\right)$. The upper bound on the range follows from taking the limit to infinity:

$$\begin{aligned}\lim_{\pi_v\to\infty}\frac{d}{d\pi_v}\mathcal{B}_v\left(\frac{1}{\pi_v}\right) &= \lim_{\pi_v\to\infty} G_v\left(\frac{1}{\pi_v}\right) - \frac{1}{\pi_v}g_v\left(\frac{1}{\pi_v}\right)\\ &= \lim_{\pi_v\to\infty}\frac{N_v K_v}{\kappa}\log\left(\frac{A_v^*+e^{\frac{\kappa}{\pi_v}}}{A_v^*+1}\right) - \frac{N_v K_v}{1+A_v^* e^{-\frac{\kappa}{\pi_v}}} \qquad (6.48)\\ &= 0\end{aligned}$$

□

Using this result, we can now prove Proposition 6.5.

Proof. When there is no penalty for violating constraint (6.14), i.e. when $\lambda = 0$, Lemma 6.5 implies that $\mathcal{L}(\lambda)$ is decreasing in π_v and hence that it is optimal to choose $\pi_v = \infty$. For a large penalty, i.e., when $\lambda < -\frac{N_v K_v}{r_v \kappa}\log\left(A+1\right)$, Lemma 6.6 implies that

$$\begin{aligned}\frac{\partial}{\partial\pi_v}\mathcal{L}(\lambda) &\geq -\frac{N_v K_v}{\kappa}\log\left(A+1\right) - r_v\lambda\\ &> 0\end{aligned}$$

Hence, the Lagrangean function increases with π_v in this case, implying that is it optimal to make π_v as small as possible: $\pi_v = 0$. In the third case, the Lagrangean problem is solved by solving $\frac{\partial\mathcal{L}(\lambda)}{\partial\pi_v} = 0$, where the derivative is calculated as:

$$\frac{\partial}{\partial \pi_v}\mathcal{L}(\lambda) = \frac{d}{d\pi_v}\mathcal{B}_v\left(\frac{1}{\pi_v}\right) - r_v\lambda \tag{6.49}$$

By Lemmas 6.5 and 6.6, we know that there is one unique value of π_v that minimizes the Lagrangean function and that this value is given by function $h_v(\lambda)$. □

6.B Upper Bounds on Screening Intervals

Among the increasing burden progression functions $f_v(s)$, those that are *convex* deserve specific attention. Though it will generally be unrealistic to assume the (disease) burden will keep increasing in the absence of screening, convex functions can provide useful approximations. For example, they are shown to adequately reflect willingness to pay for water as a function of the time without water (see, e.g., Holguín-Veras et al., 2016, 2013). One general and extremely relevant subclass of convex disease burden progression functions are those obtained when (1) expected incidence (rate of new disease cases) is an increasing function $\lambda_v(s)$ of s and (2) the expected disease burden incurred by an infected person is an increasing function $\beta(\tau - s)$ of the time since the onset of the disease $\tau - s$. The disease burden at stage τ is then calculated as in (6.8) and is proven to be convex in the following proposition.

Proposition 6.6. *The disease burden at stage of progression τ, as defined in* (6.8)*, is convex in τ if functions $\lambda_v(s)$ and $\beta(s)$ are increasing in s.*

Proof. The integral in (6.8) is the convolution of the functions $\lambda_v(s)$ and $\beta(s)$, denoted by $\lambda_v * \beta$. Taking the derivative of a convolution twice gives (Bracewell, 1965):

$$(\lambda_v * \beta)'' = \lambda_v' * \beta'.$$

Hence, this yields:

$$\frac{d^2}{d\tau^2} f_v(\tau) = \int_0^\tau \lambda_v'(s)\beta'(\tau - s)ds. \tag{6.50}$$

Because $\lambda_v' \geq 0$ and $\beta' \geq 0$, this implies that the second derivative of function (6.8) is nonnegative and hence that this function is convex.

□

Using the same reasoning, one can prove the convexity of $f_v(s)$ when one of the functions $\lambda_v(s)$ or $\beta(s)$ is convex and the other nonnegative.

Let us now return to our original planning and allocation problem, as formulated in (6.1) - (6.4). As an input for several solution methods, it can be useful to know that there exists an optimal solution for which the number of planning periods between two screening rounds in a given village v does not exceed a given (finite) upper bound. The same holds for the number of planning periods between two allocations of a team to a given cluster c. It is not difficult to come up with instances in which the optimal time interval is infinite, showing that such a finite upper bound does not exist in general. For the class of epidemics characterized by convexly increasing disease burden progression functions and a fixed screening impact, and when r_v is equal among villages (e.g., as obtained when redefining villages as sub-populations of a given size), we can establish the following upper bounds. Here, $\mathcal{R} = \left\lfloor \frac{1}{r_v} \right\rfloor$ denotes the number of villages that can be screened per team per month and $\tau = \max\left\{ \left\lceil \frac{2|V|-1}{\mathcal{R}M} \right\rceil, 3 \right\}$.

Proposition 6.7. *If $f_v(s)$ is monotonically increasing and convex, $r_v = r$ for all villages $v \in \mathcal{V}$, and $|\mathcal{C}| = 1$, then the number of planning periods between two visits of village v is bounded from above by:*

$$\hat{\tau}_v = \left\lceil \max_{w \in \mathcal{V}} \left\{ t : \frac{F_w(s_v^* + \tau - 1) - 2F_w\left(s_v^* + \frac{\tau-1}{2}\right)}{F_v(s_v^* + t) - F_v(s_v^* + 2) - F_v(s_v^* + t - 2)} = 1 \right\} \right\rceil.$$

Proof. Again, we make use of functions $g_v(s)$ and $G_v(\tau)$ as defined in (6.42) and (6.43). Suppose village v is not visited for t planning periods, and consider the case that $t \geq \tau$. There exists at least one village w that is visited at least 3 times during

the first τ of the t planning periods. Switching the second of these three visits to village v increases the burden in village w by at most

$$G_w(\tau - 1) - 2G_w\left(\frac{\tau - 1}{2}\right) \geq 0.$$

Due to the convexity of $g_v(s)$ and hence of $G_v(s)$, the largest increase is obtained when the three visits are performed at times 1, τ, and $\frac{\tau-1}{2}$. In addition, this switch decreases the burden in village v by at least

$$G_v(t) - (G_v(2) + G_v(t-2)) > 0.$$

The smallest decrease is obtained when the second visit to village w is performed at time 2. Hence the switch is always beneficial if

$$G_w\left(\tau - 1\right) - 2G_w\left(\frac{\tau - 1}{2}\right) \leq G_v(t) - (G_v(2) + G_v(t-2)) \tag{6.51}$$

Note that the right-hand side is negative when $t = 0$ and that it monotonically increases to infinity when t increases because $g_v(s)$ is convex and monotonically increasing. Hence, equality of the equation is attained for a unique value of t. The fact that we do not know village w beforehand then leads to the proposed bound. □

Proposition 6.8. *If $f_v(s)$ is monotonically increasing and convex, $r_v = r$ for all villages $v \in \mathcal{V}$, and $|\mathcal{C}| \geq 2$, then the number of planning periods between two visits of district c is bounded from above by:*

$$\hat{t}_c = \left\lceil \max_{d \in \mathcal{C}} \left\{ t : \frac{\max_{v \in V_d} \left\{ \min\{|V_d|, \mathcal{R}\} \left(F_v(s_v^* + 2\hat{\tau}_v) - 2F_v\left(s_v^* + \hat{\tau}_v\right)\right)\right\}}{\min_{v \in V_c} \left\{ \min\{|V_c|, \mathcal{R}\} \left(F_v(s_v^* + t) - F_v(s_v^* + 2) - F_v(s_v^* + t - 2)\right)\right\}} = 1 \right\} \right\rceil \tag{6.52}$$

where

$$\hat{\tau}_v = \left\lceil \max_{w \in V} \left\{ t : \frac{F_w(s_v^* + \tau - 1) - 2F_w\left(s_v^* + \frac{\tau-1}{2}\right)}{F_v(s_v^* + t) - F_v(s_v^* + 2) - F_v(s_v^* + t - 2)} = 1 \right\} \right\rceil, \tag{6.53}$$

and

$$\tau = \max\left\{ \left(\left\lceil \frac{2|V_c| - 1}{\mathcal{R}} \right\rceil - 1 \right) |\mathcal{C}| + 1, 2|\mathcal{C}| + 1 \right\}.$$

Proof. We first prove an upper bound on the time between two consecutive visits for each village in a given cluster d. Afterwards, we show that switching a team from cluster d to a different cluster c is beneficial when the time since the last visit to c exceeds a certain threshold.

Suppose that cluster c has not been visited in the last $t_c \geq \tau$ planning periods, where $\tau = \frac{(q-1)(|\mathcal{C}|-1)+1}{M}$ for some positive integer q. Then there exists at least one cluster d that is visited at least q times during the first τ of the t_c planning periods. Let us now consider a given village v in cluster d that has not been visited for $t_v \geq \tau$ planning periods and let $q = \max\left\{ \left\lceil \frac{2|\mathcal{V}_d|-1}{\mathcal{R}} \right\rceil, 3 \right\}$. Then there is at least one other village $w \in \mathcal{V}_d$ that is visited at least 3 times. Applying the "switching trick" from the proof of Proposition 6.7 yields that the number of planning periods between two visits of village v is bounded from above by $\hat{\tau}_v$ (see (6.53)).

Next, suppose that we are going to switch one of the (at least three) screening rounds in cluster d to cluster c. We know that village $v \in \mathcal{V}_d$ is visited at least three times during $2\hat{\tau}_v + 1$ planning periods. Hence, this increases the total disease burden in district d by at most

$$\max_{v \in \mathcal{V}_d} \left\{ \min\{|\mathcal{V}_d|, \mathcal{R}\} \left(G_v(2\hat{\tau}_v) - 2G_v\left(\hat{\tau}_v\right) \right) \right\},$$

and decreases the total disease burden in district c by at least

$$\min_{v \in \mathcal{V}_c} \left\{ \min\{|\mathcal{V}_c|, \mathcal{R}\} \left(G_v(t_c) - G_v(2) - G_v(t_c - 2) \right) \right\}.$$

Applying the "switching trick" again, now on a cluster level, and accounting for the fact that we do not know cluster d beforehand, we obtain the proposed bound. □

Proposition 6.9. *If $f_v(s)$ is monotonically increasing and convex, $r_v = r$ for all villages $v \in \mathcal{V}$, and $|\mathcal{C}| \geq 2$, then the number of planning periods between two visits of village $v \in \mathcal{V}_c$ is bounded from above by:*

$$\hat{\tau}_v = \left\lceil \max_{w \in \mathcal{V}} \left\{ t : \frac{F_w(s_v^* + \tau_c - 1) - 2F_w\left(s_v^* + \frac{\tau_c - 1}{2}\right)}{F_v(s_v^* + t) - F_v(s_v^* + 2) - F_v(s_v^* + t - 2)} = 1 \right\} \right\rceil,$$

where

$$\tau_c = \left\lceil \hat{\tau}_c \left\lceil \frac{2 \cdot |\mathcal{V}_c| - 1}{\mathcal{R}} - 1 \right\rceil + 1 \right\rceil,$$

and where $\hat{\tau}_c$ is as defined in (6.52).

Proof. Suppose village $v \in V_c$ is not visited for t planning periods, and consider the case that that $t \geq \tau_c$. There exists at least one village w that is visited at least 3 times during the first τ_c of the t planning periods. Applying the "switching trick" (see Proposition 6.7) between village v and some village $w \in V_c$ yields the result. □

6.C Repeated BLP Approach

An exact approach to solve problem (6.1) - (6.4) is obtained by repeatedly applying the BLP approach, as described in Algorithm 3. First, we set parameters A_{ij}^1 and A_{ij}^0 optimistically: the burden level j at the beginning of the next period is the *tightest lower bound* on the actual burden obtained when starting the current period with burden level i. Second, we we solve the BLP, determine the *actual set of burden levels* $\mathcal{F}_{va}$ encountered in each village, add this to the existing set of burden levels $\mathcal{F}_v$, update parameters A_{vij}^1 and A_{vij}^0, and resolve the BLP problem. This process is repeated until we find a solution in which the burden levels encountered mirror the actual burden levels. Optimality of this solution is guaranteed by the

fact that the quality of each other solution is represented optimistically by the choice of the parameters A^1_{vij} and A^0_{vij}. Convergence of the algorithm is implied by the observations that the algorithm always terminates when a solution is obtained for the second time and the number of possible visit patterns is finite.

Algorithm 3 *Repeated BLP Approach*

$I = 1, \mathcal{F}_v = \mathcal{F}_{v0}, \mathcal{F}_{va} = \emptyset$
while $\mathcal{F}_{va} \nsubseteq \mathcal{F}_v || I = 1$ **do**
    $\mathcal{F}_v := \mathcal{F}_v \cup \mathcal{F}_{va}$
    Update parameters A^1_{vij} and A^0_{vij} optimistically
    Solve BLP (6.24) - (6.31)
    Evaluate $\mathcal{F}_{va}$
    $I := I + 1$
end while

6.D Exact Solution Method Pricing Problem (6.39)

The following algorithm solves the pricing problem to optimality. The reason lies in the observation that the length of the shortest path found equals the reduced costs of the corresponding visit pattern. Since the length of each alternative path provides a lower bound on their respective corresponding patterns, there is no alternative pattern having lower reduced costs.

Algorithm 4 *Exact solution method for pricing problem* (6.39).

$I = 1, \mathcal{F}_v = \mathcal{F}_{v0}, \mathcal{F}_{va} = \emptyset$
while $\mathcal{F}_{va} \nsubseteq \mathcal{F}_v || I = 1$ **do**
    $\mathcal{F}_v := \mathcal{F}_v \cup \mathcal{F}_{va}$
    Build graph $\mathcal{G}(\mathcal{F}_v)$. Select arcs optimistically
    Find the shortest $s - t$ path
    Evaluate $\mathcal{F}_{va}$
    $I := I + 1$
end while

Convergence of the algorithm is again guaranteed by the observations that the algorithm terminates as soon as the same path is encountered for the second time, and since the number of possible paths is finite.

6.E Local Search Method Pricing Problem (6.39)

The idea behind the local search method is to repeatedly fix the schedule for village v for $T-k$ planning periods to and optimize the schedule for the remaining periods by enumerating all options. We first do so for $k=1$ until the algorithm converges and for $k=2$ afterwards. For a given value of parameter k, the algorithm is explained below. As with the iterated local optimization method, we run the algorithm twice, using starting solutions $\mathbf{x}_v = \mathbf{0}$ and $\mathbf{x}_v = \mathbf{1}$.

Algorithm 5 *Local search method for a given value of parameter k for pricing problem* (6.39).

while the solution changes **do**
  for $\mathcal{T}_k \subseteq \mathcal{T} : |\mathcal{T}_k| = k$ **do**
    Fix x_{vs} for all $s \in \mathcal{T} \backslash \mathcal{T}_k$
    Evaluate all options for the remaining schedule $\{x_{vt} : t \in \mathcal{T}_k\}$
    Choose the best schedule and update x_{vt}
  end for
end while

6.F Table of Notation

MSTD	
$\mathcal{T}$	set of planning periods $t \in \{1, 2, ..., T\}$
$\mathcal{V}$	set of villages v
$\mathcal{C}$	set of clusters c
$\mathcal{V}_c$	subset of villages in cluster c
y_{ct}	1 if cluster c is visited in period t; 0 otherwise
x_{vt}	1 if village v is visited in period t; 0 otherwise
r_v	fract. of available time per team per period consumed by visiting village v
M	number of mobile screening teams
$\mathcal{B}_v(\tau_v)$	average disease burden incurred in village v, given τ_v
Disease burden model	
τ_v	vector of screening intervals for village v
$f_v(s)$	disease burden in village v when the epidemic is in stage of progression s
$F_v(\tau)$	cumulative disease burden in village v between stage $s = 0$ and $s = \tau$
p	screening impact fraction
$\lambda_v(s)$	disease incidence in village v when the epidemic is in stage of progression s
$\beta(\tau - s)$	expected disease burden per person, given the time since disease onset $\tau - s$
N_v	population of village v
K_v	carrying capacity of village v (logistic model)
κ	speed of convergence parameter (logistic model)
A_v, A_{vn}, A_v^*	initial value parameter (logistic model)
Capacity-burden relationship	
π_v	screening frequency for village v
Λ	set of possible limiting values $f_v(s_\ell)$ for a given initial burden
$\underline{\tau}_{v\ell}, \overline{\tau}_{v\ell}$	bounds on interval for τ_v within which limiting value $f_v(s_\ell)$ is attained
$\bar{f}_{vn}$	avg. expected prevalence in village v between screening rounds n and $n+1$
Solution methods	
$\mathcal{F}_v$	ordered discrete set of burden levels i
$\mathcal{F}_{va}$	ordered discrete set of actual burden levels for a given solution
z_{vit}	1 if village v encounters level i at the beginning of period t; 0 otherwise
ζ_{vi1}	initial value parameter for z_{vit}
A^1_{vij}	1 if level j successes level i in village v if it is screened; 0 otherwise
A^0_{vij}	1 if level j successes level i in village v if it is not screened; 0 otherwise
$\bar{b}_{vi}$	avg. exp. burden in a period if at the period start level i is encountered
$\hat{b}_{ct}(\mathbf{x})$	sol. value improvement when choosing $y_{ct} = 1$ instead of $y_{ct} = 0$, given $\mathbf{x}$
$\hat{b}_{vt}(\mathbf{x})$	sol. value improvement when choosing $x_{vt} = 1$ instead of $x_{vt} = 0$, given $\mathbf{x}$
$\mathcal{P}$	set of screening patterns $\hat{p} \in \mathbb{B}^{\lvert\mathcal{T}\rvert}$
$\mathcal{P}_t$	subset of screening patterns $\hat{p}$ that perform a screening in month t
$x_{\hat{p}v}$	1 if pattern $\hat{p}$ is chosen for village v; 0 otherwise
$\bar{b}_{\hat{p}v}$	avg. burden incurred in village v during the planning horizon for pattern $\hat{p}$
γ_v, γ_{ct}	dual variables corresponding to constraints (6.33) and (6.34)
$rc_{\hat{p}v}$	reduced costs for pattern $\hat{p}$ and village v
rc_v^*	minimum reduced costs for village v
$\mathcal{G}$	graph used to solve pricing problem (6.39) as a shortest path problem
Cases	
part	fraction of people participating in screening rounds
$sens_{scr}$	fraction of infected people detected by the screening test
$sens_{con}$	fraction of infected people detected by the confirmation test
treat	fraction of people initiating treatment after diagnosis

Table 6.3: Key notation.

Chapter 7

Conclusions and summary

This thesis discusses how the field of humanitarian logistics could contribute to evidence-based humanitarian decision making, which is defined as "the integration of humanitarian decision makers' expertise, contextual information, and best available evidence into the decision making process for humanitarian operations". Evidence resulting from modeling and optimization studies is typically perceived to be weak, but such studies have major potential advantages over empirical research methods. We argue that, to unlock this potential, it is of key importance to consider the full spectrum of relevant objectives, choose the right objective function, analyze internal validity of models and external validity of results, and choose the right format of results and their presentation.

Next to initiating this discussion, we also contribute to the required evidence base by developing and evaluating a variety of tools, solutions, and recommendations that are related to key decisions in humanitarian operations. These analyses explicitly account for the aforementioned challenges. First, the spectrum of relevant objectives is based on extensive interactions with humanitarian organizations. Second, we use ultimate outcomes as an objective for modeling choices. Third, we employ "best available evidence" to model the link between decisions and outcomes. Fourth, we extensively analyze sensitivity of results with respect to modeling assumptions and data quality.

Critical analysis reveals several directions where future research could contribute to a stronger evidence base. First, of course, we could proceed to implementation and evaluation of our methods. For the case of sleeping sickness control, we are indeed considering empirically analyzing different planning policies. For the case of roadside clinics, experimentation is less trivial. However, a cross-sectional analysis, relating differences in health outcomes to differences in access to clinics is certainly feasible. Second, though our sensitivity analyses explore the effect of several major simplifying assumptions, there are likely other explicit or implicit assumptions for which additional analyses would improve the strength of the evidence. Third, our analyses of the trade-off between effectiveness and "costs of complexity" of tools or guidelines could be substantially deepened. For example, it would be worthwhile to investigate guidelines for the location decisions considered that use traffic count data instead of origin-destination traffic data. The latter are substantially more difficult to estimate. Moreover, to make an explicit trade-off between effectiveness and complexity, empirical analysis of the "costs of complexity" is key.

We now summarize the contributions of each chapter. *Chapter 2* discusses the potential role of advanced, optimization-based planning and routing tools in the humanitarian sector. Using expert interviews and existing literature, we show that the cost-effectiveness of such a planning system is affected by a broad mix of organizational, demand-related, and operational context factors. Several types of costs – direct costs, information losses, and costs of conflicting organizational cultures – tend to outweigh the effectiveness increase realized by advanced planning and routing in many contexts. The latter tends to be small due to triviality of planning and routing problems and external constraints that reduce the decision space. Our numerical results show that advanced approaches tend to substantially underperform when information is inaccurate, missing, or not (optimally) communicated, as is the case in many humanitarian contexts. Even if advanced systems are cost-effective, their implementation tends to have low priority due to capacity shortages and competing innovations. Our results stress the importance of basing planning system adaptations and investments in planning and routing software on thorough contextual understanding and critical cost-effectiveness analyses. Moreover, they suggest that

researchers should put more effort into investigating internal and external "validity" and incorporating costs aspects in their analyses.

Chapter 3 exemplifies the incorporation of "best available evidence" into humanitarian decision making. We consider the design problem for a network of roadside healthcare facilities, as run by NGO North Star Alliance. These facilities provide truck drivers in Africa, who carry a high burden of infectious diseases like HIV, with access to basic healthcare. Based on a review of existing evidence and expert interviews, we suggest that location decisions impact truck drivers' health in two ways: by determining their healthcare utilization and by affecting access (dis)continuities. The latter is because roadside facilities generally are the only source of healthcare for truck drivers during their trips, which often last multiple weeks. We propose means of measuring these aspects, and we incorporate them in a mixed-integer linear programming formulation for the network design problem, which we prove to be strongly NP-hard. We apply our model to the network of major transport corridors in Southern and Eastern Africa. A major insight is that substantial gains in continuity of access can be realized by making a small sacrifice in patient volume covered. This provides a strong argument for incorporating this criterion in future design decisions. Based on theoretical and computational analyses, we show that the solutions found are highly robust with respect to data inaccuracy and with respect to "suboptimality" of parameter values estimated by decision makers.

Chapter 4 builds upon Chapter 3. As North Star's network grows larger, the objective to decrease inequality in access to healthcare along the different truck routes gains importance. This chapter considers the problem of designing networks of facilities based on the objectives of maximizing healthcare utilization, ensuring continuity of access, and maximizing equity. Using the North South Corridor Network as a case study, we show that significant improvements in equity of access can be achieved with marginal loss in terms of the other objectives. More generally, solutions that are close to optimal in terms of each objective are shown to be attainable. Since the equity criterion substantially enhances computational complexity of location decisions, this chapter also investigates new models and solution methods. Specifically, we propose a novel, set-partitioning type of formulation for the problem and propose a column generation algorithm to solve it. We also propose and analyze several state-of-the-

art acceleration techniques, including dual stabilization, column pool management, and accelerated pricing. Though the facility location problem is strongly NP-hard, our algorithm yields near-optimal solutions to large, randomly generated problem instances within an acceptable amount of time.

Chapter 5 investigates evidence-based objective functions in the context of population screening for human African trypanosomiasis (HAT; sleeping sickness). Millions of people in the DRC and surrounding countries are at risk of sleeping sickness, which can severely impact health and is fatal if left untreated. Extensive screening campaigns, where mobile units screen populations at risk, have so far been the only means to control the epidemic. Important questions in this context are: Which villages should be included in the screening program? How frequently should villages be screened to ensure eradication in the long term? What is the effectiveness of planning decisions for mobile screening teams, and how can their planning be optimized? To answer such questions, it is of utmost importance to understand the relationship between (the timing of) screening rounds and the expected development of the epidemic. We introduce and test five models that describe this relationship based on historical information. We demonstrate the applicability of one of these models to develop and evaluate planning policies, presenting mathematical expressions for the relationship between participation in screening rounds, sensitivity of the diagnostic test, endemicity level in the village, and the screening frequency required to reach eradication (zero prevalence) or elimination (one case per 10 000) within a given time-frame. Applying these expressions to the Kwamouth (DRC) health zone yields estimates of the maximum screening interval that leads to eradication, the expected time to elimination, and the case detection fraction needed to reach elimination within five years, each of which has important policy implications.

Finally, *Chapter 6* considers how evidence on the relationship between planning of screening rounds and disease burden can improve planning decisions for mobile population screening teams. This chapter also analyzes the trade-off between the effectiveness of planning decisions and complexity of decision support methods. For a general class of epidemics, we present generic methods to assess the relationship between screening capacity and long-term disease burden. These methods can assist policy makers in determining how many screening teams are expected to be needed to

stabilize the epidemic or to reach eradication of the disease. Furthermore, we model the planning problem, prove it to be NP-hard and propose three flexible solution approaches – a binary linear programming approach, an iterated local optimization approach, and a column generation approach. Since simple planning rules tend to be the norm in practice, we also propose several easily implementable planning policies. Next, we numerically assess performance of our algorithms and policies, as well as sensitivity of solutions. As a case study, we consider the planning of mobile teams in DRC that screen for cases of HAT. Our analyses reveal a piece-wise linear relationship between expected HAT prevalence and screening capacity. We also show that the simple planning policies yield results that are close to those obtained by advanced approaches and that solution quality is relatively insensitive to impreciseness of input data. For our case study, we estimate that targeting screening efforts based on endemicity instead of uniformly distributing screening capacity decreases HAT prevalence by 32%. Our results suggest that the WHO's guidelines on screening frequencies could be substantially improved.

References

T. Abeku, S. De Vlas, G. Borsboom, A. Teklehaimanot, A. Kebede, D. Olana, G. Van Oortmarssen, and J. Habbema. Forecasting malaria incidence from historical morbidity patterns in epidemic-prone areas of Ethiopia: a simple seasonal adjustment method performs best. *Tropical Medicine & International Health*, 7 (10):851–857, 2002.

I. Adan and J. Resing. *Queueing Systems*. Eindhoven University of Technology, 2015.

D. Ahlburg. The economic impacts of tuberculosis. Technical report, World Health Organization, Montreal, 2000.

R. Allard. Use of time-series analysis in infectious disease surveillance. *Bulletin of the World Health Organization*, 76(4):327, 1998.

T. Andersson and P. Värbrand. Decision support tools for ambulance dispatch and relocation. *Journal of the Operational Research Society*, 58(2):195–201, 2007.

A. Anily, C. A. Glass, and R. Hassin. The scheduling of maintenance service. *Discrete Applied Mathematics*, 82(1):27–42, 1998.

Y. Apostolopoulos and S. Sönmez. Tracing the diffusion of infectious diseases in the transport sector. In Y. Apostolopoulos and S. Sönmez, editors, *Population Mobility and Infectious Disease*, pages 131–153. Springer New York, 2007.

J. Arches. Social structure, burnout, and job satisfaction. *Social Work*, 36(3):202–206, 1991.

J. N. Ares, H. De Vries, and D. Huisman. A column generation approach for locating roadside clinics in Africa based on effectiveness and equity. *European Journal of Operational Research*, 254(3):1002–1016, 2016.

M. Artzrouni and J.-P. Gouteux. Control strategies for sleeping sickness in Central Africa: a model-based approach. *Tropical Medicine & International Health*, 1(6): 753–764, 1996.

M. Artzrouni and J.-P. Gouteux. A model of Gambian sleeping sickness with open vector populations. *Mathematical Medicine and Biology*, 18(2):99–117, 2001.

I. Averbakh and O. Berman. Locating flow-capturing units on a network with multi-counting and diminishing returns to scale. *European Journal of Operational Research*, 91(3):495–506, 1996.

K. Baker Edward. Evolution of microcomputer-based vehicle routing software: case studies in the United States. *The Vehicle Routing Problem, P. Toth et D. Vigo (éditeurs), SIAM Monographs on Discrete Mathematics and Applications, Philadelphia*, pages 352–361, 2002.

R. Baldacci, E. Hadjiconstantinou, and A. Mingozzi. An exact algorithm for the capacitated vehicle routing problem based on a two-commodity network flow formulation. *Operations Research*, 52(5):723–738, 2004.

N. Banatvala and A. B. Zwi. Conflict and health: public health and humanitarian interventions: developing the evidence base. *British Medical Journal*, 321(7253): 101, 2000.

R. D. Banker, S. M. Datar, C. F. Kemerer, and D. Zweig. Software complexity and maintenance costs. *Communications of the ACM*, 36(11):81–95, 1993.

A. Bar-Noy, R. Bhatia, J. Naor, and B. Schieber. Minimizing service and operation costs of periodic scheduling. *Mathematics of Operations Research*, 27(3):518–544, 2002.

J. E. Bennett, R. Dolin, and M. J. Blaser. *Principles and practice of infectious diseases*, volume 1. Elsevier Health Sciences, 2014.

L. Berger. Northern Corridor infrastructure master plan. volume 1: Analysis and forecast of transport demand. Technical report, Paris, France, 2011.

D. P. Bertsekas. *Nonlinear programming*. Athena Scientific Belmont, 1999.

D. J. Bertsimas. A vehicle routing problem with stochastic demand. *Operations Research*, 40(3):574–585, 1992.

M. Besiou, A. Pedraza Martinez, and L. N. Van Wassenhove. The effect of earmarked funding on fleet management for relief and development. 2012.

M. Boccia, A. Sforza, and C. Sterle. Flow intercepting facility location: Problems, models and heuristics. *Journal of Mathematical Modelling and Algorithms*, 8(1): 35–79, 2009.

C. Botão, R. Z. Horth, H. Frank, B. Cummings, C. Inguane, I. Sathane, W. McFarland, H. F. Raymond, and P. W. Young. Prevalence of HIV and associated risk factors among long distance truck drivers in Inchope, Mozambique, 2012. *AIDS and Behavior*, pages 1–10, 2015.

S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

R. Bracewell. *The fourier transform and its applications*. McGraw Hill, 1965.

D. A. Bradt et al. Evidence-based decision-making in humanitarian assistance. *HPN Network Paper-Humanitarian Practice Network, Overseas Development Institute*, (67), 2009.

M. L. Brandeau. Allocating resources to control infectious diseases. In *Operations Research and Health Care*, pages 443–464. Springer, 2005.

M. L. Brandeau, G. S. Zaric, and A. Richter. Resource allocation for control of infectious diseases in multiple independent populations: beyond cost-effectiveness analysis. *Journal of health economics*, 22(4):575–598, 2003.

R. Brun, J. Blum, F. Chappuis, and C. Burri. Human African trypanosomiasis. *Lancet*, 375(9709):148–159, 2010.

J. Caldwell, P. Caldwell, J. Anarfi, K. Awusabo-Asare, J. Ntozi, I. Orubuloye, J. Marck, W. Cosford, R. Colombo, and E. Hollings. Resistances to behavioural change to reduce HIV/AIDS infection in predominantly heterosexual epidemics in third world countries. Technical report, Health Transition Centre, National Centre for Epidemiology and Population Health, Australian National University, Canberra, ACT, Australia, 1999.

T. Cardoso, M. D. Oliveira, A. Barbosa-Póvoa, and S. Nickel. An integrated approach for planning a long-term care network with uncertainty, strategic policy and equity considerations. *European Journal of Operational Research*, 247(1):321–334, 2015.

T. Cardoso, M. D. Oliveira, A. Barbosa-Póvoa, and S. Nickel. Moving towards an equitable long-term care network: A multi-objective and multi-period planning approach. *Omega*, 58:69–85, 2016.

K. Chalvet-Monfray, M. Artzrouni, J. Gouteux, P. Auger, and P. Sabatier. A two-patch model of Gambian sleeping sickness: application to vector control strategies in a village and plantations. *Acta Biotheoretica*, 46(3):207–222, 1998.

F. Checchi, J. Filipe, D. Haydon, D. Chandramohan, and F. Chappuis. Estimates of the duration of the early and late stage of gambiense sleeping sickness. *BMC Infectious Diseases*, 8(1):16, 2008.

F. Checchi, A. P. Cox, F. Chappuis, G. Priotto, D. Chandramohan, D. T. Haydon, et al. Prevalence and under-detection of gambiense human African trypanosomiasis during mass screening sessions in Uganda and Sudan. *Parasites & Vectors*, 5:157, 2012.

X. Chen, Y. Yin, X. Gong., G. Liang, W. Zhang, G. Poumerol, M. Shi, S. Wu, and G. Zhang. Prevalence of sexually transmitted infections among long-distance truck drivers in Tongling, China. *International Journal of STD & AIDS*, 17(5):304–308, 2006.

N. Christofides and S. Eilon. An algorithm for the vehicle-dispatching problem. *Journal of the Operational Research Society*, 20(3):309–318, 1969.

J. Chuma, T. Abuya, D. Memusi, E. Juma, W. Akhwale, J. Ntwiga, A. Nyandigisi, G. Tetteh, R. Shretta, A. Amin, et al. Reviewing the literature on access to prompt and effective malaria treatment in Kenya: implications for meeting the Abuja targets. *Malaria Journal*, 8(1):243, 2009.

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

C. J. Corbett and L. N. Van Wassenhove. The natural drift: What happened to operations research? *Operations Research*, 41(4):625–640, 1993.

A. Culyer and A. Wagstaff. Equity and equality in health and health care. *Journal of health economics*, 12(4):431–457, 1993.

C. F. Daganzo. The distance traveled to visit N points with a maximum of C stops per vehicle: An analytic model and an application. *Transportation Science*, 18(4): 331–350, 1984.

J. Darcy, H. Stobaugh, P. Walker, and D. Maxwell. The use of evidence in humanitarian decision making. *ACAPS operational learning paper. Feinstein International Center, Tufts University, Somerville*, 2013.

T. K. Dasaklis, C. P. Pappis, and N. P. Rachaniotis. Epidemics control and logistics operations: A review. *International Journal of Production Economics*, 139(2): 393–410, 2012.

M. Daskin and L. Dean. Location of health care facilities. In *Operations research and health care*, pages 43–76. Springer, 2004.

R. H. Davis. Waiting-time distribution of a multi-server, priority queuing system. *Operations Research*, 14(1):133–136, 1966.

S. Davis, S. Aksoy, and A. Galvani. A global sensitivity analysis for African sleeping sickness. *Parasitology*, 138(04):516–526, 2011.

L. E. De la Torre, I. S. Dolinskaya, and K. R. Smilowitz. Disaster relief routing: Integrating research and practice. *Socio-Economic Planning Sciences*, 46(1):88–97, 2012.

H. De Vries. *How to Optimise Investments in the Network of Medical Centres along the African Highways*. Master Thesis. Econometric Institute, Erasmus University Rotterdam, 2011. http://hdl.handle.net/2105/10558.

H. De Vries, J. Van de Klundert, and A. Wagelmans. Health benefits of roadside healthcare services. Technical report, Econometric Institute, Erasmus School of Economics, Januari 2014a. URL `http://repub.eur.nl/org/9724`. Report number: EI 2014-01.

H. De Vries, J. Van de Klundert, and A. Wagelmans. The roadside healthcare facility location problem. Technical report, Econometric Institute, Erasmus School of Economics, April 2014b. URL `http://repub.eur.nl/pub/51315`. Report number: EI 2014-09.

H. De Vries, A. P. Wagelmans, E. Hasker, C. Lumbala, P. Lutumba, S. J. De Vlas, and J. Van de Klundert. Forecasting human African trypanosomiasis prevalences from population screening data using continuous time models. *PLoS Computational Biology*, 12(9):e1005103, 2016.

R. Dekker. Integrating optimisation, priority setting, planning and combining of maintenance activities. *European Journal of Operational Research*, 82(2):225–240, 1995.

R. Dekker. Applications of maintenance optimization models: a review and analysis. *Reliability Engineering & System Safety*, 51(3):229–240, 1996.

S. Delany-Moretlwe, B. Bello, P. Kinross, M. Oliff, M. Chersich, I. Kleinschmidt, and H. Rees. HIV prevalence and risk in long-distance truck drivers in South Africa: a national cross-sectional survey. *International journal of STD & AIDS*, 25(6): 428–438, 2014.

S. Deo and M. Sohoni. Optimal decentralization of early infant diagnosis of HIV in resource-limited settings. *Available at SSRN 1754642*, 2013.

S. Deo, K. Rajaram, S. Rath, U. S. Karmarkar, and M. B. Goetz. Planning for HIV screening, testing, and care at the veterans health administration. *Operations Research*, 63(2):287–304, 2015.

G. Desaulniers, J. Desrosiers, and M. M. Solomon. Accelerating strategies in column generation methods for vehicle routing and crew scheduling problems. In *Essays and surveys in metaheuristics*, pages 309–324. Springer, 2002.

G. Desaulniers, J. Desrosiers, and M. M. Solomon. *Column generation*. Springer, 2005.

O. Diekmann, J. Heesterbeek, and J. A. Metz. On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4):365–382, 1990.

K. Doerner, A. Focke, and W. Gutjahr. Multicriteria tour planning for mobile healthcare facilities in a developing country. *European Journal of Operational Research*, 179(3):1078–1096, 2007.

Effective Altruism. Introduction to effective altruism, 2017. URL `https://www.effectivealtruism.org/articles/introduction-to-effective-altruism`.

M. Eftekhar and L. N. Van Wassenhove. Fleet management policies for humanitarian organizations: Beyond the utilization–residual value trade-off. *Journal of Operations Management*, 2016.

C. EM-DAT. The OFDA/CRED international disaster database. *Université catholique de Louvain*, 2017.

J. Fakcharoenphol, C. Harrelson, and S. Rao. The k-traveling repairman problem. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 655–664. Society for Industrial and Applied Mathematics, 2003.

A. Ferguson and C. Morris. Mapping transactional sex on the Northern Corridor highway in Kenya. *Health & Place*, 28(3):504–519, 2007.

E. Fevre, K. Picozzi, J. Jannin, S. Welburn, and I. Maudlin. Human African trypanosomiasis: epidemiology and control. *Advances in Parasitology*, 61:167–221, 2006.

M. Figliozzi. Planning approximations to the average length of vehicle routing problems with varying customer demands and routing constraints. *Transportation Research Record: Journal of the Transportation Research Board*, (2089):1–8, 2008.

M. Fisher. Strengthening the empirical base of operations management. *Manufacturing & Service Operations Management*, 9(4):368–382, 2007.

V. Foster and D. Benitez. The Democratic Republic of Congo's infrastructure: a continental perspective. *World Bank Policy Research Working Paper*, (5602), 2011.

P. H. Franses. *Time series models for business and economic forecasting*. Cambridge university press, 1998.

E. Frykberg. Triage: principles and practice. *Scandinavian Journal of Surgery*, 94 (4):272, 2005.

S. Funk, H. Nishiura, H. Heesterbeek, W. Edmunds, and F. Checchi. Identifying transmission cycles at the human-animal interface: the role of animal reservoirs in maintaining gambiense human African trypanosomiasis. *PLoS Comput Biol*, 9(1): e1002855, 2013.

J. Gallien, S. C. Graves, and A. Scheller-Wolf. Practice-based research in operations management: What it is, why do it, related challenges, and how to overcome them. *Manufacturing & Service Operations Management*, 18(1):5–14, 2015.

M. R. Gary and D. S. Johnson. Computers and intractability: a guide to the theory of NP-completeness, 1979.

A. Gatignon and L. Van Wassenhove. Paving the road to healthy highways: a partnership to scale up HIV & AIDS clinics in Africa. Technical report, INSEAD, Fontainebleau, 2008.

GiveWell. Our criteria for top charities, 2017. URL `http://www.givewell.org/how-we-work/criteria`.

M. Goddard and P. Smith. Equity of access to health care services: Theory and evidence from the UK. *Social science & medicine*, 53(9):1149–1162, 2001.

J. Golub, C. Mohan, G. Comstock, and R. Chaisson. Active case finding of tuberculosis: historical perspective and future prospects. *International Journal of Tuberculosis and Lung Disease*, 9(11):1183–1203, 2005.

G. Gomez, W. Venter, J. Lange, H. Rees, and C. Hankins. North-South Corridor demonstration project: Ethical and logistical challenges in the design of a demonstration study of early antiretroviral treatment for long distance truck drivers along a transport corridor through South Africa, Zimbabwe, and Zambia. *Advances in preventive medicine*, 2013.

P. Greaney, S. Pfiffner, and W. David. The Sphere handbook: Humanitarian charter and minimum standards in humanitarian response, 2011.

J. A. Griffin. *Improving health care delivery through multi-objective resource allocation.* PhD thesis, Georgia Institute of Technology, 2012.

A. Grigoriev, J. Van de Klundert, and F. Spieksma. Modeling and solving the periodic maintenance problem. *European Journal of Operational Research*, 172(3):783–797, 2006.

L. Gustavsson. Humanitarian logistics: context and challenges. *Forced Migration Review*, 18(6), 2003.

M. Hachicha, J. Hodgson, G. Laporte, and F. Semet. Heuristics for the multi-vehicle covering tour problem. *Computers & Operations Research*, 27(1):29–42, 2000.

G. H. Hardy. *Divergent series*, volume 334. American Mathematical Society, 2000.

G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities.* Cambridge University Press, 1952.

E. Hasker. Personal communication, 2016.

E. Hasker, P. Lutumba, D. Mumba, V. Lejon, P. Büscher, V. Kande, J. Muyembe, J. Menten, J. Robays, and M. Boelaert. Diagnostic accuracy and feasibility of serological tests on filter paper samples for outbreak detection of Tb gambiense human African trypanosomiasis. *Am J Trop Med Hyg*, 83(2):374, 2010.

E. Hasker, P. Lutumba, F. Chappuis, V. Kande, J. Potet, A. De Weggheleire, C. Kambo, E. Depoortere, B. Pécoul, and M. Boelaert. Human African trypanosomiasis in the Democratic Republic of the Congo: a looming emergency? *PLoS Neglected Tropical Diseases*, 6(12):e1950, 2012.

C. Heij, P. De Boer, P. Franses, T. Kloek, H. Van Dijk, et al. *Econometric methods with applications in business and economics*. Oxford University Press, 2004.

I. Heller and C. Tompkins. An extension of a theorem of dantzig's. *Linear inequalities and related systems*, 38:247–254, 1956.

H. Hethcote. The mathematics of infectious diseases. *SIAM review*, 42(4):599–653, 2000.

G. Higgs. A literature review of the use of GIS-based measures of access to health care services. *Health Services and Outcomes Research Methodology*, 5(2):119–139, 2004.

M. Hodgson. Developments in flow-based location-allocation models. *Advanced Studies in Theoretical and Applied Econometrics*, 35:119–132, 1998.

M. Hodgson, G. Laporte, and F. Semet. A covering tour model for planning mobile health care facilities in SuhumDistrict, Ghama. *Journal of Regional Science*, 38 (4):621–638, 1998.

J. Holguín-Veras, M. Jaller, L. N. Van Wassenhove, N. Pérez, and T. Wachtendorf. On the unique features of post-disaster humanitarian logistics. *Journal of Operations Management*, 30(7):494–506, 2012.

J. Holguín-Veras, N. Pérez, M. Jaller, L. N. Van Wassenhove, and F. Aros-Vera. On the appropriate objective function for post-disaster humanitarian logistics models. *Journal of Operations Management*, 31(5):262–280, 2013.

J. Holguín-Veras, J. Amaya-Leal, V. Cantillo, L. N. Van Wassenhove, F. Aros-Vera, and M. Jaller. Econometric estimation of deprivation cost functions: a contingent valuation experiment. *Journal of Operations Management*, 45:44–56, 2016.

E. W. Hook III, C. M. Richey, P. Leone, G. Bolan, C. Spalding, K. Henry, P. Clarke, M. Smith, and C. L. Celum. Delayed presentation to clinics for sexually transmitted diseases by symptomatic patients: a potential contributor to continuing STD morbidity. *Sexually transmitted diseases*, 24(8):443–448, 1997.

J. Howick, I. Chalmers, P. Glasziou, T. Greenhalgh, C. Heneghan, A. Liberati, I. Moschetti, B. Phillips, and H. Thornton. The 2011 Oxford CEBM levels of evidence (introductory document). Technical report, Oxford Centre for Evidence-Based Medicine, Oxford, 2011.

C. Hudson. AIDS in rural Africa: a paradigm for HIV-1 infection. *International journal of STD & AIDS*, 7(4):236–243, 1996.

ICRC. Annual report 2015. Technical report, 2015.

ILO. HIV/AIDS in the transport sector of Southern African countries. Technical report, International Labour Organization, Geneva, 2005.

IOM. Mobile populations and HIV/AIDS in the Southern African region: Recommendations for action. Technical report, International Organization for Migration, Geneva, 2003.

JICA. Preparatory survey for Southern Africa integrated regional transport program. Technical report, Japan International Cooperation Agency, 2010.

JICA. Data collection survey for economic and industrial development along economic corridors in Southern Africa. Technical report, Japan International Cooperation Agency, 2013.

A. D. Johnson, D. R. Thomson, S. Atwood, I. Alley, J. L. Beckerman, I. Kone, D. Diakite, H. Diallo, B. Traore, K. Traore, et al. Assessing early access to care and child survival during a health system strengthening intervention in Mali: a repeated cross sectional survey. *PloS one*, 8(12):e81304, 2013.

J. Jusot, S. De Vlas, G. Van Oortmarssen, and A. De Muynck. Apport d'un modele mathématique dans le controle d'une parasitose: cas de la trypanosomiase humaine Africaine a trypanosoma brucei gambiense. *Annales de la Société Belge de Médecine Tropicale*, 75:257–272, 1995.

N. Katoh and T. Ibaraki. Resource allocation problems. *Handbook of combinatorial optimization*, 2:159–260, 1998.

O. Kella and U. Yechiali. Waiting times in the non-preemptive priority M/M/c queue. *Stochastic Models*, 1(2):257–262, 1985.

H. Kellerer, U. Pferschy, and D. Pisinger. Knapsack problems, 2004.

R. A. Khatib, M. Selemani, G. A. Mrisho, I. M. Masanja, M. Amuri, M. H. Njozi, D. Kajungu, I. Kuepfer, S. M. Abdulla, and D. de Savigny. Access to artemisinin-based anti-malarial treatment and its related factors in rural Tanzania. *Malaria journal*, 12(155):10–1186, 2013.

T. Kiderlen, M. Conteh, S. Roll, S. Seeling, and S. Weinmann. Cross-sectional study assessing HIV-related knowledge, attitudes and behavior in the Namibian truck transport sector: Readjusting HIV prevention programs in the workplace. *Journal of Infection and Public Health*, 2015.

H. Kim and M. Stoner. Burnout and turnover intention among social workers: Effects of role stress, job autonomy and social support. *Administration in Social Work*, 32 (3):5–25, 2008.

R. Knott. Vehicle scheduling for emergency relief management: A knowledge-based approach. *Disasters*, 12(4):285–293, 1988.

P. Knox-Clarke and J. Darcy. Insufficient evidence? the quality and use of evidence in humanitarian action. *ALNAP (Active Learning Network for Accountability and Performance), London*, 2014.

J. Koskinen. Software maintenance costs. *Jyväskylä: University of Jyväskylä*, 2010.

S. Kraiselburd and P. Yadav. Supply chains and global health: An imperative for bringing operations management scholarship into action. *Production and Operations Management*, 22(2):377–381, 2013.

M. Kuby and S. Lim. The flow-refueling location problem for alternative-fuel vehicles. *Socio-Economic Planning Sciences*, 39(2):125–145, 2005.

M. Kulis, M. Chawla, A. Kozierkiewicz, and E. Subata. Truck drivers and casual sex: an inquiry into the potential spread of HIV/AIDS in the Baltic region. *World Bank Publications*, 37, 2004.

R. Lacerda, N. Gravato, W. McFarland, G. Rutherford, K. Iskrant, R. Stall, and N. Hearst. Truck drivers in Brazil: prevalence of HIV and other sexually transmitted diseases, risk behavior and potential for spread of infection. *AIDS*, 11(1): S15–S19, 1997.

U. Laukamm-Josten, B. Mwizarubi, A. Outwater, C. Mwaijonga, J. Valadez, D. Nyamwaya, R. Swai, T. Saidel, and K. Nyamuryekung'e. Preventing HIV infection through peer education and condom promotion among truck drivers and their sexual partners in Tanzania, 1990-1993. *AIDS Care*, 12(1):27–40, 2000.

J. Lee and D. Wilson. Polyhedral methods for piecewise-linear functions I: the lambda method. *Discrete applied mathematics*, 108(3):269–285, 2001.

A. Leiras, I. de Brito Jr, E. Queiroz Peres, T. Rejane Bertazzo, and H. Tsugunobu Yoshida Yoshizaki. Literature review of humanitarian logistics research: trends and challenges. *Journal of Humanitarian Logistics and Supply Chain Management*, 4 (1):95–130, 2014.

E. B. Lerner, R. B. Schwartz, P. L. Coule, E. S. Weinstein, D. C. Cone, R. C. Hunt, S. M. Sasser, J. M. Liu, N. G. Nudell, I. S. Wedmore, et al. Mass casualty triage: an

evaluation of the data and development of a proposed national guideline. *Disaster Medicine and Public Health Preparedness*, 2(S1):S25–S34, 2008.

B. Lichtenstein, E. Hook, and D. Grimley. HIV risk among long haul truckers in the USA. *Culture, Health & Sexuality*, 10(1):43–56, 2008.

M. E. Lübbecke. Column generation. *Wiley Encyclopedia of Operations Research and Management Science*, 2010.

P. Lutumba, J. Robays, C. Miaka, V. Kande, D. Mumba, P. Büscher, B. Dujardin, and M. Boelaert. Validity, cost and feasibility of the mAECT and CTC confirmation tests after diagnosis of African of sleeping sickness. *Tropical Medicine & International Health*, 11(4):470–478, 2006.

J. Lysgaard. Cvrpsep: A package of separation routines for the capacitated vehicle routing problem. 2003.

M. Malta, F. Bastos, P.-K. E.M, M. Cunha, C. Marques, and S. Strathdee. A qualitative assessment of long distance truck drivers' vulnerability to HIV/AIDS in Itajai, Southern Brazil. *AIDS care*, 18(5):489–496, 2006.

Marie Stopes International. Scaling-up excellence: Placing people at the centre of your organisation's success. Annual Fleet Forum Conference, Dublin, 2016a.

Marie Stopes International. Insight to action: MSI's research, monitoring and evaluation strategy to support scaling up excellence 2016-2020. Technical report, 2016b.

M. Marsh and D. Schilling. Equity measurement in facility location analysis: A review and framework. *European Journal of Operational Research*, 74(1):1–17, 1994.

J. Matovu and B. Ssebadduka. Sexual risk behaviours, condom use and sexually transmitted infection treatment-seeking behaviours among female sex workers and truck drivers in Uganda. *International Journal of STD & AIDS*, 23(4):267–273, 2012.

D. G. Maxwell, J. W. Parker, and H. C. Stobaugh. What drives program choice in food security crises? examining the "response analysis" question. *World Development*, 49:68–79, 2013.

J. H. McCoy and E. M. Johnson. Clinic capacity management: Planning treatment programs that incorporate adherence. *Production and Operations Management*, 23 (1):1–18, 2014.

J. H. McCoy and H. L. Lee. Using fairness models to improve equity in health delivery fleet management. *Production and Operations Management*, 23(6):965–977, 2014.

E. McKean. *The New Oxford American Dictionary*, volume 2. Oxford University Press New York, 2005.

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

J. Medlock, K. Atkins, D. Thomas, S. Aksoy, and A. Galvani. Evaluating paratransgenesis as a potential control strategy for African trypanosomiasis. *PLoS Neglected Tropical Diseases*, 7(8):e2374, 2013.

A. Meyer-Weitz, P. Reddy, H. V. den Borne, G. Kok, and J. Pietersen. Health care seeking behaviour of patients with sexually transmitted diseases: determinants of delay behaviour. *Patient education and counseling*, 41(3):263–274, 2000.

E. Mills, J. Nachega, D. Bangsberg, S. Singh, B. Rachlis, P. Wu, K. Wilson, I. Buchan, C. Gill, and C. Cooper. Adherence to HAART: a systematic review of developed and developing nation patient-reported barriers and facilitators. *PLoS Medicine*, 3(11):e438, 2006.

N. Mirasol. The output of an $M/G/\infty$ queueing system is Poisson. *Operations Research*, 11:282–284, 1963.

A. Moore and M. Richer. Re-emergence of epidemic sleeping sickness in southern Sudan. *Tropical Medicine & International Health*, 6(5):342–347, 2001.

C. Morris and A. Ferguson. Estimation of the sexual transmission of HIV in Kenya and Uganda on the trans-Africa highway: the continuing role for prevention in high risk groups. *Sexually Transmitted Infections*, 82(5):368–371, 2006.

C. Morris and A. Ferguson. Sexual and treatment-seeking behaviour for sexually transmitted infection in long-distance transport workers of East Africa. *Sexually Transmitted Infections*, 83(3):242–245, 2007.

S. Moses, E. Ngugi, J. Bradley, E. Njeru, G. Eldridge, E. Muia, J. Olenja, and F. Plummer. Health care-seeking behavior related to the transmission of sexually transmitted diseases in Kenya. *American Journal of Public Health*, 84(12):1947–1951, 1994.

M. Moura, K. Dupnik, M. Kathryn, G. Sampaio, P. Nobrega, A. Jeronimo, J. do Nascimento-Filho, R. Dantas, J. Queiroz, J. Barbosa, G. Dias, et al. Active surveillance of Hansen's disease (leprosy): Importance for case finding among extra-domiciliary contacts. *PLoS Neglected Tropical Diseases*, 7(3):e2093, 2013.

A. Mpanya, D. Hendrickx, M. Vuna, A. Kanyinda, C. Lumbala, V. Tshilombo, P. Mitashi, O. Luboya, V. Kande, M. Boelaert, et al. Should i get screened for sleeping sickness? a qualitative study in Kasai province, Democratic Republic of Congo. *PLoS Neglected Tropical Diseases*, 6(1):e1467, 2012.

S. Munro, S. Lewin, H. Smith, M. Engel, A. Fretheim, and J. Volmink. Patient adherence to tuberculosis treatment: a systematic review of qualitative research. *PLoS Medicine*, 4(7):e238, 2007.

C. J. Murray, K. F. Ortblad, C. Guinovart, S. S. Lim, T. M. Wolock, D. A. Roberts, E. A. Dansereau, N. Graetz, R. M. Barber, J. C. Brown, et al. Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990–2013: a systematic analysis for the global burden of disease study 2013. *The Lancet*, 384(9947):1005–1070, 2014.

P. Musgrove. Measurement of equity in health. *World health statistics quarterly. Rapport trimestriel de statistiques sanitaires mondiales*, 39(4):325–35, 1986.

M. Najafi, K. Eshghi, and W. Dullaert. A multi-objective robust optimization model for logistics planning in the earthquake response phase. *Transportation Research Part E: Logistics and Transportation Review*, 49(1):217–249, 2013.

Nathan Associates Inc. Corridor diagnostic study of the northern and central corridors of Eeast Africa. Technical report, 2011.

Y. Nesterov and A. Nemirovskii. *Interior-point polynomial algorithms in convex programming*, volume 13. SIAM, 1994.

R. P. Nicolai and R. Dekker. Optimal maintenance of multi-component systems: a review. In *Complex system maintenance handbook*, pages 263–286. Springer, 2008.

North Star Alliance. New ORTEC system helps North Star Alliance "optimize the world", 2011. URL `http://www.northstar-alliance.org/new-ortec-system-helps-north-star-alliance-to-optimize-the-world/`.

OCEBM. The Oxford 2011 levels of evidence, 2011.

J. Odoki and M. Anyala. Economic benefits of an efficient North-South Corridor. Technical report, Alta Innovations Limited, RTFP, 2014.

OECD. Net official development assistance (ODA) data, 2015. URL `https://data.oecd.org/oda/net-oda.htm`.

W. Ogryczak. Inequality measures and equitable approaches to location problems. *European Journal of Operational Research*, 122(2):374 – 391, 2000. ISSN 0377-2217.

W. Ogryczak. Inequality measures and equitable locations. *Annals of Operations Research*, 167(1):61–86, 2009.

M. Ortuño, P. Cristóbal, J. Ferrer, F. Martín-Campo, S. Muñoz, G. Tirado, and B. Vitoriano. Decision aid models and systems for humanitarian logistics. A survey. In *Decision aid models for disaster management and emergencies*, pages 17–44. Springer, 2013.

I. Orubuloye, P. Caldwell, and J. Caldwell. The role of high-risk occupations in the spread of AIDS: Truck drivers and itinerant market women in Nigeria. *International Family Planning Perspectives*, 19(2):43–48, 1993.

OXFAM. Annual report 2014 - 2015. Technical report, 2015.

A. Pandey, S. Benara, N. Roy, D. Sahu, M. Thomas, D. Joshi, U. Sengupta, R. Paranjape, A. Bhalla, and A. Prakash. Risk behaviour, sexually transmitted infections and HIV among long-distance truck drivers: a cross-sectional survey along national highways in India. *AIDS*, 22:S81–S90, 2008.

C. Papadimitriou and K. Steiglitz. *Combinatorial optimization: algorithms and complexity*. Courier Dover Publications, 1998.

C. Paquet, J. Castilla, D. Mbulamberi, M. Beaulieu, and A. Moren. Trypanosomiasis from trypanosoma brucei gambiense in the center of north-west Uganda. evaluation of 5 years of control (1987-1991). *Bulletin De La Societe De Pathologie Exotique*, 88(1):38–41, 1994.

A. J. Pedraza-Martinez and L. N. Van Wassenhove. Transportation and vehicle fleet management in humanitarian logistics: challenges for future research. *EURO Journal on Transportation and Logistics*, 1(1-2):185–196, 2012.

A. J. Pedraza-Martinez, S. Hasija, and L. N. Van Wassenhove. An operational mechanism design for fleet management coordination in humanitarian operations. 2010.

A. J. Pedraza-Martinez, O. Stapleton, and L. N. Van Wassenhove. Field vehicle fleet management in humanitarian operations: a case-based approach. *Journal of Operations Management*, 29(5):404–421, 2011.

J. Pépin and H. Méda. The epidemiology and control of human African trypanosomiasis. *Advances in Parasitology*, 49:71–132, 2001.

N. Pérez-Rodríguez and J. Holguín-Veras. Inventory-allocation distribution models for postdisaster humanitarian logistics with explicit consideration of deprivation costs. *Transportation Science*, 2015.

A. Pessoa, E. Uchoa, M. de Aragão, and R. Rodrigues. Exact algorithm over an arc-time-indexed formulation for parallel machine scheduling problems. *Mathematical Programming Computation*, 2(3-4):259–290, 2010. ISSN 1867-2949.

A. Peterson. *The origin-destination matrix estimation problem: analysis and computations*. PhD thesis, Institutionen för teknik och naturvetenskap, 2007.

M. Piot. A simulation model of case finding and treatment in tuberculosis control programmes. Technical report, 1967.

N. Pollock, R. Williams, and R. Procter. Fitting standard software packages to non-standard organizations: the 'biography' of an enterprise-wide system. *Technology Analysis & Strategic Management*, 15(3):317–332, 2003.

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

G. Ramjee and E. Gouws. Prevalence of HIV among truck drivers visiting sex workers in KwaZulu-Natal, South Africa. *Sexually Transmitted Diseases*, 29(1):44–49, 2002.

R. Ranganathan and V. Foster. The SADC's infrastructure: A regional perspective. *World Bank Policy Research Working Paper Series, Vol*, 2011.

A. Roa, K. Misra, and K. Verma. A national multicentric study in india to determine STD incidence amongst intercity truck drivers. In *Program and Abstracts of the XI International AIDS Conference*, page 161, Vancouver, Brittish Columbia, 1997.

J. Robays, M. M. C. Bilengue, P. V. d. Stuyft, and M. Boelaert. The effectiveness of active population screening and treatment for sleeping sickness control in the Democratic Republic of Congo. *Tropical Medicine & International Health*, 9(5): 542–550, 2004.

D. W. Robertson, R. Bedell, J. V. Lavery, and R. Upshur. What kind of evidence do we need to justify humanitarian medical aid? *The Lancet*, 360(9329):330, 2002.

K. S. Rock, S. J. Torr, C. Lumbala, and M. J. Keeling. Quantitative evaluation of the strategy to eliminate human African trypanosomiasis in the Democratic Republic of Congo. *Parasites & Vectors*, 8(1):1–13, 2015.

D. Rogers. A general model for the African trypanosomiases. *Parasitology*, 97(01): 193–212, 1988.

D. Rogers. Satellites, space, time and the African trypanosomiases. *Advances in Parasitology*, 47:129–171, 2000.

D. M. Rousseau. Is there such a thing as "evidence-based management"? *Academy of management review*, 31(2):256–269, 2006.

D. M. Rousseau. *The Oxford handbook of evidence-based management*. Oxford University Press, 2012.

G. Royston. Meeting global health challenges through operational research and management science. *Bulletin of the World Health Organization*, 89(9):683–688, 2011.

J. A. Ruiz, P. P. Simarro, and T. Josenando. Control of human African trypanosomiasis in the Quicama focus, Angola. *Bulletin of the World Health Organization*, 80(9):738–745, 2002.

J. Sachs and P. Malaney. The economic and social burden of malaria. *Nature*, 415 (6872):680–685, 2002.

D. L. Sackett, W. M. Rosenberg, J. Gray, R. B. Haynes, and W. S. Richardson. Evidence based medicine: what it is and what it isn't. *BMJ: British Medical Journal*, 312(7023):71, 1996.

SADC. Regional infrastructure development master plan: Transport sector plan. Technical report, Southern African Development Community, 2012.

Save the Children. Super: Save-university partnerships for education research, 2017. URL `http://www.savethechildren.org`.

A. Schrijver. *Theory of linear and integer programming*. John Wiley & Sons, 1998.

J. Shiffman. Donor funding priorities for communicable disease control in the developing world. *Health Policy and Planning*, 21(6):411–420, 2006.

P. Simarro, F. Sima, M. Mir, M. Mateo, and J. Roche. Control of human African trypanosomiasis in luba in equatorial Guinea: evaluation of three methods. *Bulletin of the World Health Organization*, 69(4):451–457, 1990.

P. Simarro, G. Cecchi, J. Franco, M. Paone, A. D. J. Ruiz-Postigo, E. Fèvre, R. Mattioli, and J. Jannin. Estimating and mapping the population at risk of sleeping sickness. *PLoS Neglected Tropical Diseases*, 6(10):e1859, 2012.

P. Simarro, J. Franco, A. Diarra, J. Ruiz Postigo, and J. Jannin. Diversity of human African trypanosomiasis epidemiological settings requires fine-tuning control strategies to facilitate disease elimination. *Research and Reports in Tropical Medicine*, 4:1–6, 2013.

P. P. Simarro, G. Cecchi, M. Paone, J. R. Franco, A. Diarra, J. A. Ruiz, E. M. Fèvre, F. Courtin, R. C. Mattioli, and J. G. Jannin. The atlas of human African trypanosomiasis: a contribution to global mapping of neglected tropical diseases. *International Journal of Health Geographics*, 9(1):57–75, 2010.

P. P. Simarro, A. Diarra, J. A. R. Postigo, J. R. Franco, and J. G. Jannin. The human African trypanosomiasis control and surveillance programme of the World Health Organization 2000–2009: the way forward. *PLoS Neglected Tropical Diseases*, 5 (2):e1007, 2011.

H. Smith, P. Harper, C. Potts, and A. Thyle. Planning sustainable community health schemes in rural areas of developing countries. *European Journal of Operational Research*, 193(3):768–777, 2009.

A. Solomon, J. Doucette, E. Garland, and T. McGinn. Healthcare and the long haul: Long distance truck drivers-a medically underserved population. *American Journal of Industrial Medicine*, 46(5):463–471, 2004.

O. Stapleton, A. Pedraza Martinez, and L. N. Van Wassenhove. Last mile vehicle supply chain in the international federation of red cross and red crescent societies. 2009.

B. Starfield, L. Shi, and J. Macinko. Contribution of primary care to health systems and health. *Milbank Quarterly*, 83(3):457–502, 2005.

C. M. Stone and N. Chitnis. Implications of heterogeneous biting exposure and animal hosts on trypanosomiasis brucei gambiense transmission and control. *PLoS Comput Biol*, 11(10):e1004514, 2015.

J. Stover and L. Bollinger. The economic impact of AIDS. *Washington, DC: Futures Group*, 1999.

D. Stratford, T. Ellerbrock, J. Akins, and H. Hall. Highway cowboys, old hands and Christian truckers: risk behavior for human immunodeficiency virus infection among long-haul truckers in Florida. *Social Science & Medicine*, 50(5):737–749, 2000.

F. Tanser. Methodology for optimising location of new primary health care facilities in rural communities: a case study in KwaZulu-Natal, South Africa. *Journal of Epidemiology and Community health*, 60(10):846–850, 2006.

F. Tanser, B. Gijsbertsen, and K. Herbst. 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, 2006.

S. Teravaninthorn and G. Raballand. *Transport prices and costs in Africa: a review of the main international corridors*. World Bank Publications, 2009.

P. Truc, V. Lejon, E. Magnus, V. Jamonneau, A. Nangouma, D. Verloo, L. Penchenier, and P. Büscher. Evaluation of the micro-CATT, CATT/Trypanosoma brucei gambiense, and LATEX/T. b. gambiense methods for serodiagnosis and surveillance of human African trypanosomiasis in West and Central Africa. *Bulletin of the World Health Organization*, 80(11):882–886, 2002.

TTCA. The Northern Corridor transport observatory report issue no.7. Technical report, Northern Corridor Transit and Transport Coordination Authority, 2015.

UNHCR. Evaluation of the utilization and management of UNHCR's light vehicle fleet. Technical report, 2006.

United Nations. Constitution of the world health organization. Technical report, United Nations, New York, 1946.

H. Üster and P. Kewcharoenwong. Strategic design and analysis of a relay network in truckload transportation. *Transportation Science*, 45(4):505–523, 2011.

J. Van de Klundert. Healthcare analytics: Big data, little evidence. In *Optimization Challenges in Complex, Networked and Risky Systems*, pages 307–328. INFORMS, 2016.

S. Van Nieuwenhove, V. Betu-Ku-Mesu, P. Diabakana, J. Declercq, and C. Bilenge. Sleeping sickness resurgence in the DRC: the past decade. *Tropical Medicine & International Health*, 6(5):335–341, 2001.

L. N. Van Wassenhove. Humanitarian aid logistics: supply chain management in high gear†. *Journal of the Operational Research Society*, 57(5):475–489, 2006.

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

P. Vinck. *World disasters report: Focus on technology and the future of humanitarian action*. International Federation of Red Cross and Red Crescent Societies, 2013.

M. Vledder, J. Friedman, M. Sjöblom, T. Brown, and P. Yadav. Optimal supply chain structure for distributing essential drugs in low income countries: results from a randomized experiment. *Ross School of Business Paper No. 1269*, 2015.

J. Volmink and P. Garner. Directly observed therapy for treating tuberculosis. *Cochrane Database Syst Rev*, 4, 2007.

R. J. Waldman. Prioritising health care in complex emergencies. *The Lancet*, 357 (9266):1427–1429, 2001.

K. Walshe and T. G. Rundall. Evidence-based management: from theory to practice in health care. *Milbank Quarterly*, 79(3):429–457, 2001.

H. Waters. Measuring equity in access to health care. *Social Science & Medicine*, 51 (4):599–612, 2000.

W. Wei and C. Liu. On a periodic maintenance problem. *Operations Research Letters*, 2(2):90–93, 1983.

S. C. Welburn, D. H. Molyneux, and I. Maudlin. Beyond tsetse–implications for research and control of human African trypanosomiasis epidemics. *Trends in Parasitology*, 32(3):230–241, 2016.

L. White, H. Smith, and C. Currie. OR in developing countries: a review. *European Journal of Operational Research*, 208(1):1–11, 2011.

WHO. International development assistance and health: report of the working group 6 of the commission on macroeconomics and health. Technical report, World Health Organization, Geneva, 2002.

WHO. Report of a WHO meeting on elimination of African trypanosomiasis (trypanosoma brucei gambiense). Technical report, 2013.

WHO. Human african trypanosomiasis (sleeping sickness), factsheet 259, May 2015a. URL `www.who.int/mediacentre/factsheets/fs259/en/`.

WHO. Control and surveillance of human African trypanosomiasis. Technical Report 984, 2015b.

WHO. Global health observatory data repository, 2016. URL `http://apps.who.int/ghodata`.

WHO/UNICEF. Africa malaria report 2003. Technical report, World Health Organization, Geneva, 2003.

A. Williams and R. Cookson. Equity in health. *Handbook of health economics*, 1: 1863–1910, 2000.

B. Williams, R. Dransfield, and R. Brightwell. The control of tsetse flies in relation to fly movement and trapping efficiency. *Journal of Applied Ecology*, pages 163–179, 1992.

D. Wilson. Corridors of Hope in Southern Africa: HIV prevention needs and opportunities in four border towns. Technical report, Family Health International, 2005.

D. B. Winters, J. T. Lindley, S. Topping, and L. T. Lindley. The hidden financial costs of ERP software. *Managerial Finance*, 34(2):78–90, 2008.

W. Wong, S. Tam, and P. Leung. Cross-border truck drivers in Hong Kong: Their psychological health, sexual dysfunctions and sexual risk behaviors. *Journal of Travel Medicine*, 14(1):20–30, 2007.

World Bank. Poverty and shared prosperity 2016: Taking on inequality. Technical report, World Bank, Washington D.C., 2016.

World Vision International. Evidence and learning, 2017. URL `http://www.wvi.org/impact`.

P. Yadav. Improving public health in developing countries through operations research. *Wiley Encyclopedia of Operations Research and Management Science*, 2010.

H. Young. *Equity: in theory and practice*. Princeton University Press, 1995.

G. S. Zaric and M. L. Brandeau. Resource allocation for epidemic control over short time horizons. *Mathematical Biosciences*, 171(1):33–58, 2001.

Nederlandse Samenvatting (Summary in Dutch)

Dit proefschrift bespreekt hoe het vakgebied van de humanitaire logistiek kan bijdragen aan op bewijs gefundeerd (evidence-based) beslissen bij het bieden van humanitaire hulp. Dit definiëren we als "het integreren van de expertise van de beslisser, contextuele informatie, en het best beschikbare bewijs in het beslisproces voor humanitaire operaties". Bewijs dat voortkomt uit modelleerstudies wordt over het algemeen als vrij zwak gezien. Toch hebben dit soort studies grote potentiële voordelen ten opzichte van empirische studies. Om dit potentieel te benutten is het van groot belang dat modellen rekening houden met het volledige spectrum aan relevante doelen, dat de doelfunctie goed wordt gefundeerd, dat interne validiteit van modellen en externe validiteit van resultaten wordt geanalyseerd, en dat resultaten in het juiste format worden gepresenteerd.

Naast het initiëren van deze discussie dragen we bij aan het genereren van het vereiste bewijs zelf. Dit doen we door methoden, oplossingen, en aanbevelingen te ontwikkelen voor een aantal belangrijke typen beslissingen in de humanitaire sector. Onze analyses houden expliciet rekening met de bovengenoemde uitdagingen. Ten eerste is het spectrum aan relevante doelen bepaald op basis van nauwe samenwerking met de humanitaire organisaties zelf. Ten tweede nemen we uiteindelijke uitkomsten als een basis voor modelleerkeuzes. Ten derde gebruiken we het "best beschikbare bewijs" om de link tussen beslissingen en uitkomsten te modelleren. Tot slot maken

we uitgebreide analyses van de gevoeligheid van resultaten ten opzichte van aannames en kwaliteit van data.

Kritische reflectie legt verschillende richtingen bloot waarin toekomstig onderzoek de kwaliteit van het bewijs zou kunnen verhogen. Ten eerste kunnen we natuurlijk overgaan tot implementatie en evaluatie van onze methoden. Binnen onze studie naar slaapziektebestrijding overwegen we inderdaad om verschillende planningsmethoden empirisch te testen. Experimenteren ligt minder voor de hand binnen onze studie naar het plaatsen van klinieken langs Afrikaanse snelwegen. Wat in deze context wel mogelijk is, is een cross-sectionele studie, waarbij verschillen in toegang tot zorg worden gerelateerd aan verschillen in gezondheidsuitkomsten. Ten tweede, hoewel onze gevoeligheidsanalyses het effect van verscheidene versimpelende aannames hebben getest, zijn er wellicht andere impliciete of expliciete aannames waarvoor extra analyses het bewijs zouden kunnen versterken. Ten derde zouden onze analyses van de trade-off tussen effectiviteit en de "kosten van complexiteit" kunnen worden verdiept. Zo zou men bijvoorbeeld kunnen kijken naar richtlijnen voor locatiebeslissingen die gebruik maken van verkeerstellingen in plaats van tellingen per vertrekpunt-bestemming combinatie. De laatstgenoemde zijn veel lastiger te verkrijgen dan de eerste. Om de trade-off tussen effectiviteit en complexiteit expliciet te maken zou daarnaast een empirische studie naar de kosten van complexiteit kunnen worden gedaan.

In wat volgt zullen we de bijdragen van individuele hoofdstukken samenvatten. *Hoofdstuk 2* analyseert de potentiële rol van geavanceerde optimalisatiemethoden voor plannings- en routeringsbeslissingen in de humanitaire sector. Door middel van interviews met experts en de bestaande literatuur laten we zien dat de kosteneffectiviteit van een planningssysteem afhangt van een groot aantal organisationele, vraag gerelateerde, en operationele factoren. De eventuele effectiviteitsstijging die wordt gerealiseerd door geavanceerde plannings- en routeringssystemen weegt vaak niet op tegen de verschillende typen kosten van deze systemen: directe kosten, mogelijk informatie verlies, en mogelijke kosten van conflicterende organisationele culturen. De stijging in effectiviteit is bovendien vaak klein door trivialiteit van plannings- en routeringsproblemen en door externe restricties die de beslissingsruimte verkleinen. Onze numerieke resultaten laten daarnaast zien dat geavanceerde tools erg slecht kunnen

presteren wanneer informatie onnauwkeurig of onvolledig is of wanneer het suboptimaal wordt gecommuniceerd, zoals vaak het geval is in humanitaire situaties. Zelfs wanneer geavanceerde systemen kosteneffectief zijn heeft hun implementatie vaak een lage prioriteit wegens capaciteitsgebrek en andere gewenste innovaties. Onze resultaten laten hiermee zien dat het erg belangrijk is om aanpassingen aan planning- en routeringssystemen te baseren op een grondige analyse van context en kosteneffectiviteit. Bovendien suggereren ze dat onderzoekers meer nadruk zouden moeten leggen op het analyseren van interne en externe "validiteit" en op het meenemen van kostenaspecten in hun analyses.

Hoofdstuk 3 laat zien hoe het "best beschikbare bewijs" de basis kan vormen voor beslissingen in de humanitaire sector. We bekijken het vraagstuk hoe een netwerk van klinieken langs Afrikaanse snelwegen te ontwerpen, zoals het netwerk dat wordt gerund door organisatie North Star Alliance. Deze klinieken voorzien Afrikaanse vrachtwagenchauffeurs van basale medische hulp. Toegang tot zorg is erg belangrijk voor deze populatie, aangezien chauffeurs een hoge ziektelast van HIV en andere infectieziekten dragen en ook bijdragen aan de verspreiding ervan. Op basis van een review van beschikbaar bewijs en interviews met experts motiveren we dat locatiebeslissingen de gezondheid van vrachtwagenchauffeurs op twee manieren beïnvloeden: Ze bepalen de mate waarin chauffeurs de voorzieningen *benutten* en ze bepalen de "gaten" tussen klinieken die langs een truckroute worden gepasseerd en daarmee de (dis)*continuïteit van toegang* tot zorg. Klinieken langs de weg vormen over het algemeen namelijk de enige bron van gezondheidszorg gedurende de ritten, welke vaak weken duren. We ontwikkelen modellen om deze aspecten te kunnen kwantificeren en integreren die in een mixed-integer linear programming formulering voor het netwerkontwikkelingsprobleem. Ook laten we zien dat dit probleem sterk NP-moeilijk is. We passen ons model toe op het netwerk van transport corridors in het Zuiden en Oosten van Afrika. Een van de belangrijkste inzichten die volgt uit deze analyse is dat je, zonder veel in te leveren op patiënten aantallen, een substantiële verbetering kunt realiseren in continuïteit van toegang. Dit vormt een sterk argument om dit criterium mee te nemen bij toekomstige locatiebeslissingen. Bovendien laten we met theoretische en numerieke analyses zien dat de kwaliteit van de locatiekeuzes vrij ongevoelig is voor onnauwkeurigheid in data en modelparameters.

Hoofdstuk 4 bouwt voort op Hoofdstuk 3. Terwijl North Star's netwerk van klinieken uitbreidt, wordt het doel om grote verschillen tussen de mate van toegang tot zorg tussen de verschillende transportstromen te minimaliseren steeds belangrijker. Dit hoofdstuk bekijkt het probleem hoe een netwerk van klinieken te ontwerpen op basis van 3 doelen: Het maximaliseren van de benutting van gezondheidszorg, het verbeteren van de continuïteit van toegang, en het minimaliseren van ongelijkheid. Op basis van een studie naar het North South Corridor netwerk laten we zien dat een substantiële vermindering van ongelijkheid kan worden bereikt door een kleine opoffering in patiënt volume en continuïteit van toegang. Ook laten we zien dat er netwerken bestaan waarbij ieder van de drie doelen vrijwel maximaal wordt behaald. Omdat het (on)gelijkheidscriterium het locatieprobleem substantieel complexer maakt, bekijkt dit hoofdstuk ook alternatieve modellen en oplosmethoden. Specifiek stellen we een soort set-partitioning probleemformulering voor en ontwikkelen we een kolomgeneratiemethode om dit op te lossen. We analyseren ook verschillende technieken om deze methode te versnellen, waaronder dual pricing, column pool management, en accelerated pricing. Hoewel het locatieprobleem sterk NP-moeilijk is, levert ons algoritme binnen afzienbare tijd vrijwel optimale oplossingen voor grote, willekeurig gegenereerde probleem instanties.

Hoofdstuk 5 onderzoekt op bewijs gebaseerde doelfuncties voor de bestrijding van de zogenaamde slaapziekte (human African trypanosomiasis; HAT). Miljoenen mensen in de DRC en omliggende landen lopen het risico om besmet te raken met deze ziekte, welke een enorme gezondheidsimpact kan hebben en fataal is wanneer het niet snel genoeg wordt behandeld. Uitgebreide screeningsprogramma's, waarbij mobiele teams de bevolking uit dorpen in risicogebieden screenen, zijn tot nu toe de enige manier om de epidemie te bedwingen. Belangrijke vraagstukken in deze context zijn: Welke dorpen moeten worden bezocht? Hoe frequent moeten dorpen worden bezocht om de ziekte uiteindelijk te kunnen uitroeien? Wat is de effectiviteit van planningsbeslissingen voor de mobiele teams, en hoe kunnen deze beslissingen worden geoptimaliseerd? Om dit soort vragen te kunnen beantwoorden is het van cruciaal belang te begrijpen hoe (het tijdstip van) screening verband houdt met de verwachte ontwikkeling van de epidemie. Wij ontwikkelen en analyseren vijf modellen die deze relatie beschrijven op basis van historische informatie. Ook laten we zien

hoe je zo'n model kunt inzetten om beleid voor planningsbeslissingen te ontwikkelen en te analyseren. Dit doen we specifiek door wiskundige formules te presenteren die laten zien hoe de screeningsfrequentie benodigd voor eliminatie (minder dan een geval per 10000) of uitroeiing afhangt van participatie in screeningsactiviteiten, de gevoeligheid van de diagnostische test, en de endemiciteit. Deze formules passen we toe op de Kwamouth (DRC) regio, wat resulteert in schattingen van de maximale screeningsinterval die leidt tot uitroeiing, de verwachte tijd tot eliminatie, en de fractie van de ziektegevallen die je moet detecteren per bezoek om eliminatie binnen vijf jaar te bereiken. Ieder van deze resultaten heeft belangrijke implicaties voor beleid.

Tot slot bekijkt *Hoofdstuk 6* hoe bewijs over de relatie tussen planning van populatiescreening en ziektelast kan worden ingezet om planningsbeslissingen voor mobiele screeningsteams te verbeteren. Ook analyseert dit hoofdstuk de trade-off tussen de effectiviteit van planningsbeslissingen en de complexiteit van tools om zulke beslissingen te ondersteunen. Voor een omvangrijke klasse van epidemieën presenteren we generieke methoden om de relatie tussen screeningscapaciteit en ziektelast in kaart te brengen. Deze methoden kunnen beleidsmakers helpen bij het bepalen hoeveel teams nodig zijn om de epidemie te stabiliseren of uit te roeien. Ook modelleren we het planningsprobleem voor de mobiele teams, bewijzen we dat dit probleem sterk NP-moeilijk is, en stellen we drie flexibele oplosmethoden voor. Deze maken respectievelijk gebruik van binary linear programming, iterated local optimization, en kolomgeneratie. Omdat simpele methoden vaak de norm zijn in de praktijk stellen we ook verscheidene simpele planningsregels voor. Vervolgens analyseren we de kwaliteit van onze algoritmes en planningsregels en de gevoeligheid van oplossingen. Hierbij gebruiken we de planning van mobiele slaapziektebestrijdingsteams in de DRC als testcase. Onze analyses onthullen een stuksgewijs lineaire relatie tussen de verwachte slaapziekteprevalentie en screeningscapaciteit. We laten ook zien dat een van de simpele planningsregels vrijwel even goed presteert als geavanceerde methoden en dat de kwaliteit van oplossingen vrij ongevoelig is voor onnauwkeurigheid in de benodigde data. Voor onze testcase schatten wij dat een op endemiciteit gebaseerde inzet van screeningscapaciteit in plaats van een uniforme inzet de prevalentie met 32% kan

verminderen. Onze resultaten suggereren hiermee dat de richtlijnen van de WHO waarschijnlijk substantieel kunnen worden verbeterd.

Curriculum Vitae

As an academic researcher, Harwin de Vries focuses on the logistics behind humanitarian aid delivery. Based on close cooperation with humanitarian organizations, Harwin analyzes how humanitarian logistics systems impact beneficiaries and how they could be improved. One major branch of his research considers the cost-effectiveness of decision support tools in this sector. A second branch considers the design of such tools themselves. Harwin won several prizes with his research, including the INFORMS Healthcare Best Student Paper Award 2015, and published several academic papers in scientific journals, including Omega, PLoS Computational Biology, and European Journal of Operational Research.

Harwin holds an MSc degree, cum laude, in Operations Research & Quantitative Logistics from the Erasmus University Rotterdam. In 2013, he started his PhD candidacy at the Erasmus Research Institute of Management at Erasmus University Rotterdam. His teaching experience includes combinatorial optimization, linear programming, simulation, numerical methods, and statistics. Harwin presently works as a Post-Doctoral Research Fellow at the INSEAD Social Innovation Centre, specifically with the Humanitarian Research Group.

Portfolio

Publications

Publications in journals:

De Vries, H., Duijzer, L.E., (2017). Incorporating Driving Range Variability in Network Design for Refueling Facilities. *Omega, 69*, 102-114

De Vries, H., Wagelmans, A. P., Hasker, E., Lumbala, C., Lutumba, P., De Vlas, S. J., Van de Klundert, J. (2016). Forecasting human African trypanosomiasis prevalences from population screening data using continuous time models. *PLoS Computational Biology, 12*(9), e1005103.

Nunez Ares, J., De Vries, H., Huisman, D. (2016). A column generation approach for locating roadside clinics in Africa based on effectiveness and equity. *European Journal of Operational Research, 254*(3), 1002-1016.

De Vries, H., Carrasco-Gallego, R., Farenhorst-Yuan, T., Dekker, R. (2014). Prioritizing replenishments of the piece picking area. *European Journal of Operational Research, 236(*1), 126-134.

Working papers and reports:

De Vries, H., Van de Klundert, J.J., Wagelmans, A.P.M. The Roadside Healthcare Facility Location Problem. Submitted to *Production and Operations Management journal (POM).*

De Vries, H., Van de Klundert, J.J., Wagelmans, A.P.M. Optimizing Population Screening for Infectious Diseases. *In preparation for journal submission.*

De Vries, H., Van Wassenhove, L., Evidence-Based Vehicle Planning for Humanitarian Field Operations. Submitted to *Manufacturing & Service Operations Management (MSOM).*

De Vries, H, Van de Klundert, J.J, Wagelmans, A.P.M. (2013). Health Benefits of Roadside Healthcare Services. *Econometric Institute Research Papers EI 2014-01.*

Teaching (tutorial lecturer)

Linear Programming, Erasmus School of Economics, Econometrics and Operations Research, 2011-2012 & 2014-2015.
Combinatorial Optimization, Erasmus School of Economics, Econometrics and Operations Research, 2015-2016.
Introduction to statistics, Erasmus School of Economics, Econometrics and Operations Research, 2015-2016.
Simulation, Erasmus School of Economics, Econometrics and Operations Research, 2011-2012, 2012-2013 & 2013-2014.
Numerical Methods, Erasmus School of Economics, Econometrics and Operations Research, 2011-2012 .

PhD courses

Randomized Algorithms
Algorithms and Complexity
Markov Decision Processes
Integer Programming Methods
Operations Research and Health Care
Stochastic Programming
Networks and Polyhedra
Convex Analysis for Optimization
Multi-Class Queues and Stochastic Networks
Networks and Semidefinite Programming
Spring school on Column Generation
Publishing Strategy
Scientific Integrity
English (CPE certificate)

Conferences attended

LNMB conference 2013, Lunteren, The Netherlands
POMS conference 2013, Denver, USA
EURO 2013, Rome, Italy
OR 2013, Rotterdam, the Netherlands
LNMB conference 2014, Lunteren, The Netherlands
MSOM conference 2014, Seattle, USE
LNMB conference 2015, Lunteren, The Netherlands
INFORMS Healthcare 2015, Nashville, USA
HHL conference 2015, Johannesburg, South Africa
LNMB conference 2016, Lunteren, The Netherlands
Fleet Forum 2016, Dublin, Ireland
POMS conference 2016, Orlando, USA

The ERIM PhD Series

The ERIM PhD Series contains PhD dissertations in the field of Research in Management defended at Erasmus University Rotterdam and supervised by senior researchers affiliated to the Erasmus Research Institute of Management (ERIM). All dissertations in the ERIM PhD Series are available in full text through the ERIM Electronic Series Portal: http://repub.eur.nl/pub. ERIM is the joint research institute of the Rotterdam School of Management (RSM) and the Erasmus School of Economics at the Erasmus University Rotterdam (EUR).

Dissertations in the last five years

Abbink, E.J., *Crew Management in Passenger Rail Transport*, Promotors: Prof. L.G. Kroon & Prof. A.P.M. Wagelmans, EPS-2014-325-LIS, http://repub.eur.nl/pub/76927

Acar, O.A., *Crowdsourcing for Innovation: Unpacking Motivational, Knowledge and Relational Mechanisms of Innovative Behavior in Crowdsourcing Platforms*, Promotor: Prof. J.C.M. van den Ende, EPS-2014-321-LIS, http://repub.eur.nl/pub/76076

Akemu, O., *Corporate Responses to Social Issues: Essays in Social Entrepreneurship and Corporate Social Responsibility*, Promotors: Prof. G.M. Whiteman & Dr S.P. Kennedy, EPS-2017-392-ORG, https://repub.eur.nl/pub/95768

Akin Ates, M., *Purchasing and Supply Management at the Purchase Category Level: Strategy, structure and performance*, Promotors: Prof. J.Y.F. Wynstra & Dr E.M. van Raaij, EPS-2014-300-LIS, http://repub.eur.nl/pub/50283

Akpinar, E., *Consumer Information Sharing*, Promotor: Prof. A. Smidts, EPS-2013-297-MKT, http://repub.eur.nl/pub/50140

Alexander, L., *People, Politics, and Innovation: A Process Perspective*, Promotors: Prof. H.G. Barkema & Prof. D.L. van Knippenberg, EPS-2014-331-S&E, http://repub.eur.nl/pub/77209

Alexiou, A. *Management of Emerging Technologies and the Learning Organization: Lessons from the Cloud and Serious Games Technology*, Promotors: Prof. S.J. Magala, Prof. M.C. Schippers and Dr I. Oshri, EPS-2016-404-ORG, http://repub.eur.nl/pub/93818

Almeida e Santos Nogueira, R.J. de, *Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty*, Promotors: Prof. U. Kaymak & Prof. J.M.C. Sousa, EPS-2014-310-LIS, http://repub.eur.nl/pub/51560

Bannouh, K., *Measuring and Forecasting Financial Market Volatility using High-frequency Data*, Promotor: Prof. D.J.C. van Dijk, EPS-2013-273-F&A, http://repub.eur.nl/pub/38240

Ben-Menahem, S.M., *Strategic Timing and Proactiveness of Organizations*, Promotors: Prof. H.W. Volberda & Prof. F.A.J. van den Bosch, EPS-2013-278-S&E, http://repub.eur.nl/pub/39128

Benschop, N, *Biases in Project Escalation: Names, frames & construal levels*, Promotors: Prof. K.I.M. Rhode, Prof. H.R. Commandeur, Prof. M. Keil & Dr A.L.P. Nuijten, EPS-2015-375-S&E, http://repub.eur.nl/pub/79408

Berg, W.E. van den, *Understanding Salesforce Behavior using Genetic Association Studies*, Promotor: Prof. W.J.M.I. Verbeke, EPS-2014-311-MKT, http://repub.eur.nl/pub/51440

Beusichem, H.C. van, *Firms and Financial Markets: Empirical Studies on the Informational Value of Dividends, Governance and Financial Reporting*, Promotors: Prof. A. de Jong & Dr G. Westerhuis, EPS-2016-378-F&A, http://repub.eur.nl/pub/93079

Bliek, R. de, *Empirical Studies on the Economic Impact of Trust*, Promotor: Prof. J. Veenman & Prof. Ph.H.B.F. Franses, EPS-2015-324-ORG, http://repub.eur.nl/pub/78159

Boons, M., *Working Together Alone in the Online Crowd: The Effects of Social Motivations and Individual Knowledge Backgrounds on the Participation and Performance of Members of Online Crowdsourcing Platforms*, Promotors: Prof. H.G. Barkema & Dr D.A. Stam, EPS-2014-306-S&E, http://repub.eur.nl/pub/50711

Bouman, P., *Passengers, Crowding and Complexity: Models for Passenger Oriented Public Transport*, Prof. L.G. Kroon, Prof. A. Schöbel & Prof. P.H.M. Vervest, EPS-2017-420-LIS, https://repub.eur.nl/

Brazys, J., *Aggregated Marcoeconomic News and Price Discovery*, Promotor: Prof. W.F.C. Verschoor, EPS-2015-351-F&A, http://repub.eur.nl/pub/78243

Byington, E., *Exploring Coworker Relationships: Antecedents and Dimensions of Interpersonal Fit, Coworker Satisfaction, and Relational Models*, Promotor: Prof. D.L. van Knippenberg, EPS-2013-292-ORG, http://repub.eur.nl/pub/41508

Cancurtaran, P., *Essays on Accelerated Product Development*, Promotors: Prof. F. Langerak & Prof. G.H. van Bruggen, EPS-2014-317-MKT, http://repub.eur.nl/pub/76074

Caron, E.A.M., *Explanation of Exceptional Values in Multi-dimensional Business Databases*, Promotors: Prof. H.A.M. Daniels & Prof. G.W.J. Hendrikse, EPS-2013-296-LIS, http://repub.eur.nl/pub/50005

Carvalho, L. de, *Knowledge Locations in Cities: Emergence and Development Dynamics*, Promotor: Prof. L. Berg, EPS-2013-274-S&E, http://repub.eur.nl/pub/38449

Chammas, G., *Portfolio concentration*, Promotor: Prof. J. Spronk, EPS-2017-410-F&E, https://repub.eur.nl/pub/94975

Cranenburgh, K.C. van, *Money or Ethics: Multinational corporations and religious organisations operating in an era of corporate responsibility*, Prof. L.C.P.M. Meijs, Prof. R.J.M. van Tulder & Dr D. Arenas, EPS-2016-385-ORG, http://repub.eur.nl/pub/93104

Consiglio, I., *Others: Essays on Interpersonal and Consumer Behavior*, Promotor: Prof. S.M.J. van Osselaer, EPS-2016-366-MKT, http://repub.eur.nl/pub/79820

Cox, R.H.G.M., *To Own, To Finance, and To Insure - Residential Real Estate Revealed*, Promotor: Prof. D. Brounen, EPS-2013-290-F&A, http://repub.eur.nl/pub/40964

Darnihamedani, P. *Individual Characteristics, Contextual Factors and Entrepreneurial Behavior*, Promotors: Prof. A.R. Thurik & S.J.A. Hessels, EPS-2016-360-S&E, http://repub.eur.nl/pub/93280

Dennerlein, T. *Empowering Leadership and Employees' Achievement Motivations: the Role of Self-Efficacy and Goal Orientations in the Empowering Leadership Process*, Promotors: Prof. D.L. van Knippenberg & Dr J. Dietz, EPS-2017-414-ORG, https://repub.eur.nl/pub/98438

Deng, W., *Social Capital and Diversification of Cooperatives*, Promotor: Prof. G.W.J. Hendrikse, EPS-2015-341-ORG, http://repub.eur.nl/pub/77449

Depecik, B.E., *Revitalizing brands and brand: Essays on Brand and Brand Portfolio Management Strategies*, Promotors: Prof. G.H. van Bruggen, Dr Y.M. van Everdingen and Dr M.B. Ataman, EPS-2016-406-MKT, http://repub.eur.nl/pub/93507

Dollevoet, T.A.B., *Delay Management and Dispatching in Railways*, Promotor: Prof. A.P.M. Wagelmans, EPS-2013-272-LIS, http://repub.eur.nl/pub/38241

Duyvesteyn, J.G. *Empirical Studies on Sovereign Fixed Income Markets*, Promotors: Prof. P. Verwijmeren & Prof. M.P.E. Martens, EPS-2015-361-F&A, https://repub.eur.nl/pub/79033

Duursema, H., *Strategic Leadership: Moving Beyond the Leader-Follower Dyad*, Promotor: Prof. R.J.M. van Tulder, EPS-2013-279-ORG, http://repub.eur.nl/pub/39129

Elemes, A, *Studies on Determinants and Consequences of Financial Reporting Quality*, Promotor: Prof. E. Peek, EPS-2015-354-F&A, https://repub.eur.nl/pub/79037

Ellen, S. ter, *Measurement, Dynamics, and Implications of Heterogeneous Beliefs in Financial Markets*, Promotor: Prof. W.F.C. Verschoor, EPS-2015-343-F&A, http://repub.eur.nl/pub/78191

Erlemann, C., *Gender and Leadership Aspiration: The Impact of the Organizational Environment*, Promotor: Prof. D.L. van Knippenberg, EPS-2016-376-ORG, http://repub.eur.nl/pub/79409

Eskenazi, P.I., *The Accountable Animal*, Promotor: Prof. F.G.H. Hartmann, EPS-2015-355-F&A, http://repub.eur.nl/pub/78300

Evangelidis, I., *Preference Construction under Prominence*, Promotor: Prof. S.M.J. van Osselaer, EPS-2015-340-MKT, http://repub.eur.nl/pub/78202

Faber, N., *Structuring Warehouse Management*, Promotors: Prof. M.B.M. de Koster & Prof. A. Smidts, EPS-2015-336-LIS, http://repub.eur.nl/pub/78603

Feng, Y., *The Effectiveness of Corporate Governance Mechanisms and Leadership Structure: Impacts on strategic change and firm performance*, Promotors: Prof. F.A.J. van den Bosch, Prof. H.W. Volberda & Dr J.S. Sidhu, EPS-2017-389-S&E, https://repub.eur.nl/pub/98470

Fernald, K., *The Waves of Biotechnological Innovation in Medicine: Interfirm Cooperation Effects and a Venture Capital Perspective*, Promotors: Prof. E. Claassen, Prof. H.P.G. Pennings & Prof. H.R. Commandeur, EPS-2015-371-S&E, http://hdl.handle.net/1765/79120

Fisch, C.O., *Patents and trademarks: Motivations, antecedents, and value in industrialized and emerging markets*, Promotors: Prof. J.H. Block, Prof. H.P.G. Pennings & Prof. A.R. Thurik, EPS-2016-397-S&E, http://repub.eur.nl/pub/94036

Fliers, P.T., *Essays on Financing and Performance: The role of firms, banks and board*, Promotor: Prof. A. de Jong & Prof. P.G.J. Roosenboom, EPS-2016-388-F&A, http://repub.eur.nl/pub/93019

Fourne, S.P., *Managing Organizational Tensions: A Multi-Level Perspective on Exploration, Exploitation and Ambidexterity*, Promotors: Prof. J.J.P. Jansen & Prof. S.J. Magala, EPS-2014-318-S&E, http://repub.eur.nl/pub/76075

Gaast, J.P. van der, *Stochastic Models for Order Picking Systems*, Promotors: Prof. M.B.M de Koster & Prof. I.J.B.F. Adan, EPS-2016-398-LIS, http://repub.eur.nl/pub/93222

Giurge, L., *A Test of Time; A temporal and dynamic approach to power and ethics*, Promotors: Prof. M.H. van Dijke & Prof. D. De Cremer, EPS-2017-412-ORG, https://repub.eur.nl/

Glorie, K.M., *Clearing Barter Exchange Markets: Kidney Exchange and Beyond*, Promotors: Prof. A.P.M. Wagelmans & Prof. J.J. van de Klundert, EPS-2014-329-LIS, http://repub.eur.nl/pub/77183

Hekimoglu, M., *Spare Parts Management of Aging Capital Products*, Promotor: Prof. R. Dekker, EPS-2015-368-LIS, http://repub.eur.nl/pub/79092

Heyde Fernandes, D. von der, *The Functions and Dysfunctions of Reminders*, Promotor: Prof. S.M.J. van Osselaer, EPS-2013-295-MKT, http://repub.eur.nl/pub/41514

Hogenboom, A.C., *Sentiment Analysis of Text Guided by Semantics and Structure*, Promotors: Prof. U. Kaymak & Prof. F.M.G. de Jong, EPS-2015-369-LIS, http://repub.eur.nl/pub/79034

Hogenboom, F.P., *Automated Detection of Financial Events in News Text*, Promotors: Prof. U. Kaymak & Prof. F.M.G. de Jong, EPS-2014-326-LIS, http://repub.eur.nl/pub/77237

Hollen, R.M.A., *Exploratory Studies into Strategies to Enhance Innovation-Driven International Competitiveness in a Port Context: Toward Ambidextrous Ports*, Promotors: Prof. F.A.J. Van Den Bosch & Prof. H.W.Volberda, EPS-2015-372-S&E, http://repub.eur.nl/pub/78881

Hout, D.H. van, *Measuring Meaningful Differences: Sensory Testing Based Decision Making in an Industrial Context; Applications of Signal Detection Theory and Thurstonian Modelling*, Promotors: Prof. P.J.F. Groenen & Prof. G.B. Dijksterhuis, EPS-2014-304-MKT, http://repub.eur.nl/pub/50387

Houwelingen, G.G. van, *Something To Rely On*, Promotors: Prof. D. de Cremer & Prof. M.H. van Dijke, EPS-2014-335-ORG, http://repub.eur.nl/pub/77320

Hurk, E. van der, *Passengers, Information, and Disruptions*, Promotors: Prof. L.G. Kroon & Prof. P.H.M. Vervest, EPS-2015-345-LIS, http://repub.eur.nl/pub/78275

Iseger, P. den, *Fourier and Laplace Transform Inversion with Applications in Finance*, Promotor: Prof. R. Dekker, EPS-2014-322-LIS, http://repub.eur.nl/pub/76954

Jaarsveld, W.L. van, *Maintenance Centered Service Parts Inventory Control*, Promotor: Prof. R. Dekker, EPS-2013-288-LIS, http://repub.eur.nl/pub/39933

Khanagha, S., *Dynamic Capabilities for Managing Emerging Technologies*, Promotor: Prof. H.W. Volberda, EPS-2014-339-S&E, http://repub.eur.nl/pub/77319

Khattab, J., *Make Minorities Great Again: a contribution to workplace equity by identifying and addressing constraints and privileges*, Prof. D.L. van Knippenberg & Dr A. Nederveen Pieterse, EPS-2017-421-ORG, https://repub.eur.nl/pub/99311

Kil, J., *Acquisitions Through a Behavioral and Real Options Lens*, Promotor: Prof. H.T.J. Smit, EPS-2013-298-F&A, http://repub.eur.nl/pub/50142

Klooster, E. van't, *Travel to Learn: the Influence of Cultural Distance on Competence Development in Educational Travel*, Promotors: Prof. F.M. Go & Prof. P.J. van Baalen, EPS-2014-312-MKT, http://repub.eur.nl/pub/51462

Koendjbiharie, S.R., *The Information-Based View on Business Network Performance: Revealing the Performance of Interorganizational Networks*, Promotors: Prof. H.W.G.M. van Heck & Prof. P.H.M. Vervest, EPS-2014-315-LIS, http://repub.eur.nl/pub/51751

Koning, M., *The Financial Reporting Environment: The Role of the Media, Regulators and Auditors*, Promotors: Prof. G.M.H. Mertens & Prof. P.G.J. Roosenboom, EPS-2014-330-F&A, http://repub.eur.nl/pub/77154

Konter, D.J., *Crossing Borders with HRM: An Inquiry of the Influence of Contextual Differences in the Adoption and Effectiveness of HRM*, Promotors: Prof. J. Paauwe & Dr L.H. Hoeksema, EPS-2014-305-ORG, http://repub.eur.nl/pub/50388

Korkmaz, E., *Bridging Models and Business: Understanding Heterogeneity in Hidden Drivers of Customer Purchase Behavior*, Promotors: Prof. S.L. van de Velde & Prof. D. Fok, EPS-2014-316-LIS, http://repub.eur.nl/pub/76008

Krämer, R., *A license to mine? Community organizing against multinational corporations*, Promotors: Prof. R.J.M. van Tulder & Prof. G.M. Whiteman, EPS-2016-383-ORG, http://repub.eur.nl/pub/94072

Kroezen, J.J., *The Renewal of Mature Industries: An Examination of the Revival of the Dutch Beer Brewing Industry*, Promotor: Prof. P.P.M.A.R. Heugens, EPS-2014-333-S&E, http://repub.eur.nl/pub/77042

Kysucky, V., *Access to Finance in a Cros-Country Context*, Promotor: Prof. L. Norden, EPS-2015-350-F&A, http://repub.eur.nl/pub/78225

Lee, C.I.S.G, *Big Data in Management Research: Exploring New Avenues*, Promotors: Prof. S.J. Magala & Dr W.A. Felps, EPS-2016-365-ORG, http://repub.eur.nl/pub/79818

Legault-Tremblay, P.O., *Corporate Governance During Market Transition: Heterogeneous responses to Institution Tensions in China*, Promotor: Prof. B. Krug, EPS-2015-362-ORG, http://repub.eur.nl/pub/78649

Lenoir, A.S. *Are You Talking to Me? Addressing Consumers in a Globalised World*, Promotors: Prof. S. Puntoni & Prof. S.M.J. van Osselaer, EPS-2015-363-MKT, http://repub.eur.nl/pub/79036

Leunissen, J.M., *All Apologies: On the Willingness of Perpetrators to Apologize*, Promotors: Prof. D. de Cremer & Dr M. van Dijke, EPS-2014-301-ORG, http://repub.eur.nl/pub/50318

Li, D., *Supply Chain Contracting for After-sales Service and Product Support*, Promotor: Prof. M.B.M. de Koster, EPS-2015-347-LIS, http://repub.eur.nl/pub/78526

Li, Z., *Irrationality: What, Why and How*, Promotors: Prof. H. Bleichrodt, Prof. P.P. Wakker, & Prof. K.I.M. Rohde, EPS-2014-338-MKT, http://repub.eur.nl/pub/77205

Liu, N., *Behavioral Biases in Interpersonal Contexts*, Supervisors: Prof. A. Baillon & Prof. H. Bleichrodt, EPS-2017-408-MKT, https://repub.eur.nl/pub/95487

Liang, Q.X., *Governance, CEO Identity, and Quality Provision of Farmer Cooperatives*, Promotor: Prof. G.W.J. Hendrikse, EPS-2013-281-ORG, http://repub.eur.nl/pub/39253

Liket, K., *Why 'Doing Good' is not Good Enough: Essays on Social Impact Measurement*, Promotors: Prof. H.R. Commandeur & Dr K.E.H. Maas, EPS-2014-307-STR, http://repub.eur.nl/pub/51130

Loos, M.J.H.M. van der, *Molecular Genetics and Hormones: New Frontiers in Entrepreneurship Research*, Promotors: Prof. A.R. Thurik, Prof. P.J.F. Groenen, & Prof. A. Hofman, EPS-2013-287-S&E, http://repub.eur.nl/pub/40081

Lu, Y., *Data-Driven Decision Making in Auction Markets*, Promotors: Prof. H.W.G.M. van Heck & Prof. W. Ketter, EPS-2014-314-LIS, http://repub.eur.nl/pub/51543

Ma, Y., *The Use of Advanced Transportation Monitoring Data for Official Statistics*, Promotors: Prof. L.G. Kroon and Dr J. van Dalen, EPS-2016-391-LIS, http://repub.eur.nl/pub/80174

Manders, B., *Implementation and Impact of ISO 9001*, Promotor: Prof. K. Blind, EPS-2014-337-LIS, http://repub.eur.nl/pub/77412

Mell, J.N., *Connecting Minds: On The Role of Metaknowledge in Knowledge Coordination*, Promotor: Prof. D.L. van Knippenberg, EPS-2015-359-ORG, http://hdl.handle.net/1765/78951

Meulen,van der, D., *The Distance Dilemma: the effect of flexible working practices on performance in the digital workplace*, Promotors: Prof. H.W.G.M. van Heck & Prof. P.J. van Baalen, EPS-2016-403-LIS, http://repub.eur.nl/pub/94033

Micheli, M.R., *Business Model Innovation: A Journey across Managers' Attention and Inter-Organizational Networks*, Promotor: Prof. J.J.P. Jansen, EPS-2015-344-S&E, http://repub.eur.nl/pub/78241

Milea, V., *News Analytics for Financial Decision Support*, Promotor: Prof. U. Kaymak, EPS-2013-275-LIS, http://repub.eur.nl/pub/38673

Moniz, A, *Textual Analysis of Intangible Information*, Promotors: Prof. C.B.M. van Riel, Prof. F.M.G de Jong & Dr G.A.J.M. Berens, EPS-2016-393-ORG, http://repub.eur.nl/pub/93001

Mulder, J. *Network design and robust scheduling in liner shipping*, Promotors: Prof. R. Dekker & Dr W.L. van Jaarsveld, EPS-2016-384-LIS, http://repub.eur.nl/pub/80258

Naumovska, I., *Socially Situated Financial Markets: A Neo-Behavioral Perspective on Firms, Investors and Practices*, Promotors: Prof. P.P.M.A.R. Heugens & Prof. A. de Jong, EPS-2014-319-S&E, http://repub.eur.nl/pub/76084

Neerijnen, P., *The Adaptive Organization: the socio-cognitive antecedents of ambidexterity and individual exploration*, Promotors: Prof. J.J.P. Jansen, P.P.M.A.R. Heugens & Dr T.J.M. Mom, EPS-2016-358-S&E, http://repub.eur.nl/pub/93274

Okbay, A., *Essays on Genetics and the Social Sciences*, Promotors: Prof. A.R. Thurik, Prof. Ph.D. Koellinger & Prof. P.J.F. Groenen, EPS-2017-413-S&E, https://repub.eur.nl/pub/95489

Oord, J.A. van, *Essays on Momentum Strategies in Finance*, Promotor: Prof. H.K. van Dijk, EPS-2016-380-F&A, http://repub.eur.nl/pub/80036

Peng, X., *Innovation, Member Sorting, and Evaluation of Agricultural Cooperatives*, Promotor: Prof. G.W.J. Hendriks, EPS-2017-409-ORG, https://repub.eur.nl/pub/94976

Pennings, C.L.P., *Advancements in Demand Forecasting: Methods and Behavior*, Promotors: Prof. L.G. Kroon, Prof. H.W.G.M. van Heck & Dr J. van Dalen, EPS-2016-400-LIS, http://repub.eur.nl/pub/94039

Peters, M., *Machine Learning Algorithms for Smart Electricity Markets*, Promotor: Prof. W. Ketter, EPS-2014-332-LIS, http://repub.eur.nl/pub/77413

Pocock, M., *Status Inequalities in Business Exchange Relations in Luxury Markets*, Promotors: Prof. C.B.M. van Riel & Dr G.A.J.M. Berens, EPS-2017-346-ORG, https://repub.eur.nl/pub/98647

Porck, J., *No Team is an Island: An Integrative View of Strategic Consensus between Groups*, Promotors: Prof. P.J.F. Groenen & Prof. D.L. van Knippenberg, EPS-2013-299-ORG, http://repub.eur.nl/pub/50141

Pozharliev, R., *Social Neuromarketing: The role of social context in measuring advertising effectiveness*, Promotors: Prof. W.J.M.I. Verbeke & Prof. J.W. van Strien, EPS-2017-402-MKT, https://repub.eur.nl/pub/95528

Pronker, E.S., *Innovation Paradox in Vaccine Target Selection*, Promotors: Prof. H.J.H.M. Claassen & Prof. H.R. Commandeur, EPS-2013-282-S&E, http://repub.eur.nl/pub/39654

Protzner, S. *Mind the gap between demand and supply: A behavioral perspective on demand forecasting*, Promotors: Prof. S.L. van de Velde & Dr L. Rook, EPS-2015-364-LIS, *http*://repub.eur.nl/pub/79355

Pruijssers, J.K., *An Organizational Perspective on Auditor Conduct*, Promotors: Prof. J. van Oosterhout & Prof. P.P.M.A.R. Heugens, EPS-2015-342-S&E, http://repub.eur.nl/pub/78192

Retel Helmrich, M.J., *Green Lot-Sizing*, Promotor: Prof. A.P.M. Wagelmans, EPS-2013-291-LIS, http://repub.eur.nl/pub/41330

Rietdijk, W.J.R. *The Use of Cognitive Factors for Explaining Entrepreneurship*, Promotors: Prof. A.R. Thurik & Prof. I.H.A. Franken, EPS-2015-356-S&E, http://repub.eur.nl/pub/79817

Rietveld, N., *Essays on the Intersection of Economics and Biology*, Promotors: Prof. A.R. Thurik, Prof. Ph.D. Koellinger, Prof. P.J.F. Groenen, & Prof. A. Hofman, EPS-2014-320-S&E, http://repub.eur.nl/pub/76907

Rösch, D. *Market Efficiency and Liquidity*, Promotor: Prof. M.A. van Dijk, EPS-2015-353-F&A, http://repub.eur.nl/pub/79121

Roza, L., *Employee Engagement in Corporate Social Responsibility: A collection of essays*, Promotor: L.C.P.M. Meijs, EPS-2016-396-ORG, http://repub.eur.nl/pub/93254

Rubbaniy, G., *Investment Behaviour of Institutional Investors*, Promotor: Prof. W.F.C. Verschoor, EPS-2013-284-F&A, http://repub.eur.nl/pub/40068

Schoonees, P. *Methods for Modelling Response Styles*, Promotor: Prof.dr P.J.F. Groenen, EPS-2015-348-MKT, http://repub.eur.nl/pub/79327

Schouten, M.E., *The Ups and Downs of Hierarchy: the causes and consequences of hierarchy struggles and positional loss*, Promotors; Prof. D.L. van Knippenberg & Dr L.L. Greer, EPS-2016-386-ORG, http://repub.eur.nl/pub/80059

Shahzad, K., *Credit Rating Agencies, Financial Regulations and the Capital Markets*, Promotor: Prof. G.M.H. Mertens, EPS-2013-283-F&A, http://repub.eur.nl/pub/39655

Smit, J. *Unlocking Business Model Innovation: A look through the keyhole at the inner workings of Business Model Innovation*, Promotor: H.G. Barkema, EPS-2016-399-S&E, http://repub.eur.nl/pub/93211

Sousa, M.J.C. de, *Servant Leadership to the Test: New Perspectives and Insights*, Promotors: Prof. D.L. van Knippenberg & Dr D. van Dierendonck, EPS-2014-313-ORG, http://repub.eur.nl/pub/51537

Spliet, R., *Vehicle Routing with Uncertain Demand*, Promotor: Prof. R. Dekker, EPS-2013-293-LIS, http://repub.eur.nl/pub/41513

Staadt, J.L., *Leading Public Housing Organisation in a Problematic Situation: A Critical Soft Systems Methodology Approach*, Promotor: Prof. S.J. Magala, EPS-2014-308-ORG, http://repub.eur.nl/pub/50712

Stallen, M., *Social Context Effects on Decision-Making: A Neurobiological Approach*, Promotor: Prof. A. Smidts, EPS-2013-285-MKT, http://repub.eur.nl/pub/39931

Szatmari, B., *We are (all) the champions: The effect of status in the implementation of innovations*, Promotors: Prof J.C.M & Dr D. Deichmann, EPS-2016-401-LIS, http://repub.eur.nl/pub/94633

Tarakci, M., *Behavioral Strategy: Strategic Consensus, Power and Networks*, Promotors: Prof. D.L. van Knippenberg & Prof. P.J.F. Groenen, EPS-2013-280-ORG, http://repub.eur.nl/pub/39130

Tuijl, E. van, *Upgrading across Organisational and Geographical Configurations*, Promotor: Prof. L. van den Berg, EPS-2015-349-S&E, http://repub.eur.nl/pub/78224

Tuncdogan, A., *Decision Making and Behavioral Strategy: The Role of Regulatory Focus in Corporate Innovation Processes*, Promotors: Prof. F.A.J. van den Bosch, Prof. H.W. Volberda, & Prof. T.J.M. Mom, EPS-2014-334-S&E, http://repub.eur.nl/pub/76978

Uijl, S. den, *The Emergence of De-facto Standards*, Promotor: Prof. K. Blind, EPS-2014-328-LIS, http://repub.eur.nl/pub/77382

Vagias, D., *Liquidity, Investors and International Capital Markets*, Promotor: Prof. M.A. van Dijk, EPS-2013-294-F&A, http://repub.eur.nl/pub/41511

Valogianni, K. *Sustainable Electric Vehicle Management using Coordinated Machine Learning*, Promotors: Prof. H.W.G.M. van Heck & Prof. W. Ketter, EPS-2016-387-LIS, http://repub.eur.nl/pub/93018

Vandic, D., *Intelligent Information Systems for Web Product Search*, Promotors: Prof. U. Kaymak & Dr Frasincar, EPS-2017-405-LIS, https://repub.eur.nl/pub/95490

Veelenturf, L.P., *Disruption Management in Passenger Railways: Models for Timetable, Rolling Stock and Crew Rescheduling*, Promotor: Prof. L.G. Kroon, EPS-2014-327-LIS, http://repub.eur.nl/pub/77155

Venus, M., *Demystifying Visionary Leadership: In search of the essence of effective vision communication*, Promotor: Prof. D.L. van Knippenberg, EPS-2013-289-ORG, http://repub.eur.nl/pub/40079

Vermeer, W., *Propagation in Networks:The impact of information processing at the actor level on system-wide propagation dynamics*, Promotor: Prof. P.H.M.Vervest, EPS-2015-373-LIS, http://repub.eur.nl/pub/79325

Versluis, I., *Prevention of the Portion Size Effect*, Promotors: Prof. Ph.H.B.F. Franses & Dr E.K. Papies, EPS-2016-382-MKT, http://repub.eur.nl/pub/79880

Vishwanathan, P., *Governing for Stakeholders: How Organizations May Create or Destroy Value for their Stakeholders*, Promotors: Prof. J. van Oosterhout & Prof. L.C.P.M. Meijs, EPS-2016-377-ORG, http://repub.eur.nl/pub/93016

Visser, V.A., *Leader Affect and Leadership Effectiveness: How leader affective displays influence follower outcomes*, Promotor: Prof. D.L. van Knippenberg, EPS-2013-286-ORG, http://repub.eur.nl/pub/40076

Vlaming, R. de., *Linear Mixed Models in Statistical Genetics*, Prof. A.R. Thurik, Prof. P.J.F. Groenen & Prof. Ph.D. Koellinger, EPS-2017-416-S&E, https://repub.eur.nl/pub/100428

Vries, J. de, *Behavioral Operations in Logistics*, Promotors: Prof. M.B.M de Koster & Prof. D.A. Stam, EPS-2015-374-LIS, http://repub.eur.nl/pub/79705

Wagenaar, J.C., *Practice Oriented Algorithmic Disruption Management in Passenger Railways*, Prof. L.G. Kroon & Prof. A.P.M. Wagelmans, EPS-2016-390-LIS, http://repub.eur.nl/pub/93177

Wang, P., *Innovations, status, and networks*, Promotors: Prof. J.J.P. Jansen & Dr V.J.A. van de Vrande, EPS-2016-381-S&E, http://repub.eur.nl/pub/93176

Wang, R. *Corporate Environmentalism in China*, Promotors: Prof. P.P.M.A.R Heugens & Dr F.Wijen, EPS-2017-417-S&E, https://repub.eur.nl/pub/99987

Wang, T., *Essays in Banking and Corporate Finance*, Promotors: Prof. L. Norden & Prof. P.G.J. Roosenboom, EPS-2015-352-F&A, http://repub.eur.nl/pub/78301

Wang, Y., *Corporate Reputation Management: Reaching Out to Financial Stakeholders*, Promotor: Prof. C.B.M. van Riel, EPS-2013-271-ORG, http://repub.eur.nl/pub/38675

Weenen, T.C., *On the Origin and Development of the Medical Nutrition Industry*, Promotors: Prof. H.R. Commandeur & Prof. H.J.H.M. Claassen, EPS-2014-309-S&E, http://repub.eur.nl/pub/51134

Wessels, C., *Flexible Working Practices: How Employees Can Reap the Benefits for Engagement and Performance*, Promotors: Prof. H.W.G.M. van Heck, Prof. P.J. van Baalen & Prof. M.C. Schippers, EPS-2017-418-LIS, https://repub.eur.nl/

Wolfswinkel, M., *Corporate Governance, Firm Risk and Shareholder Value*, Promotor: Prof. A. de Jong, EPS-2013-277-F&A, http://repub.eur.nl/pub/39127

Yang, S., *Information Aggregation Efficiency of Prediction Markets*, Promotor: Prof. H.W.G.M. van Heck, EPS-2014-323-LIS, http://repub.eur.nl/pub/77184

Ypsilantis, P., *The Design, Planning and Execution of Sustainable Intermodal Port-hinterland Transport Networks*, Promotors: Prof. R.A. Zuidwijk & Prof. L.G. Kroon, EPS-2016-395-LIS, http://repub.eur.nl/pub/94375

Yuferova, D. *Price Discovery, Liquidity Provision, and Low-Latency Trading*, Promotors: Prof. M.A. van Dijk & Dr D.G.J. Bongaerts, EPS-2016-379-F&A, http://repub.eur.nl/pub/93017

Zaerpour, N., *Efficient Management of Compact Storage Systems*, Promotor: Prof. M.B.M. de Koster, EPS-2013-276-LIS, http://repub.eur.nl/pub/38766

Zuber, F.B., *Looking at the Others: Studies on (un)ethical behavior and social relationships in organizations*, Promotor: Prof. S.P. Kaptein, EPS-2016-394-ORG, http://repub.eur.nl/pub/94388

Humanitarian crises like the Syrian war, Ebola, the earthquake in Haiti, the Indian Ocean tsunami, and the ongoing HIV epidemic prompt substantial demands for humanitarian aid. Logistics plays a key role in aid delivery and represents a major cost category for humanitarian organizations.

Optimizing logistics has long been at the core of operations research: the discipline that explores the use of advanced analytical methods to improve decision making. The commercial sector has substantially benefited from such methods. This thesis discusses whether and how such methods can also guide policy and decision making in the humanitarian sector. This is done through detailed analyses of three case studies. The first investigates suitability of advanced planning and routing tools. Next, we investigate decision support methods for designing networks of roadside HIV clinics. The third case study concerns the deployment of mobile teams that screen for infectious disease outbreaks.

Optimization tools come with assumptions about objectives to be reached and about their link with the decisions to be supported. In humanitarian logistics, safeguarding adequacy of these assumptions is complex but crucial. Throughout our case-studies, we explore how best available evidence can be used to link decisions to objectives, i.e., how to enable evidence-based optimization in humanitarian logistics.

ERiM

The Erasmus Research Institute of Management (ERIM) is the Research School (Onderzoekschool) in the field of management of the Erasmus University Rotterdam. The founding participants of ERIM are Rotterdam School of Management (RSM), and the Erasmus School of Economics (ESE). ERIM was founded in 1999 and is officially accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW). The research undertaken by ERIM is focused on the management of the firm in its environment, its intra- and interfirm relations, and its business processes in their interdependent connections.

The objective of ERIM is to carry out first rate research in management, and to offer an advanced doctoral programme in Research in Management. Within ERIM, over three hundred senior researchers and PhD candidates are active in the different research programmes. From a variety of academic backgrounds and expertises, the ERIM community is united in striving for excellence and working at the forefront of creating new business knowledge.

ERiM

ERIM PhD Series
Research in Management

Erasmus University Rotterdam (EUR)
Erasmus Research Institute of Management
Mandeville (T) Building
Burgemeester Oudlaan 50
3062 PA Rotterdam, The Netherlands

P.O. Box 1738
3000 DR Rotterdam, The Netherlands
T +31 10 408 1182
E info@erim.eur.nl
W www.erim.eur.nl