The Utility of Merging Mechanistic and Geospatial Models for Malaria Intervention Planning and Policy Guidance

Amelia Bertozzi-Villa

A thesis presented for the degree of
Doctor of Philosophy in Clinical Medicine

Exeter College
University of Oxford
Trinity Term 2021

To Oxford

Contents

List of Abbreviations v

List of Figures vi

List of Tables xii

Abstract 1

Introduction 3

I Assessing Intervention Impact 11

1 An Archetypes Approach to Intervention Modeling 13

1.1 Introduction 13

1.2 Methods 19

1.2.1 Covariate Selection 20

1.2.2 Dimensionality Reduction 22

1.2.3 Clustering 23

1.2.4 Selecting Representative Sites 25

1.2.5 Selecting and Configuring a Malaria Model 26

1.2.6 Recreating Maps 28

1.2.7 Details on the EMOD Software 29

1.3 Results . . . 32
1.3.1 Covariate Scaling, Dimensionality Reduction, and Clustering . 32
1.3.2 Archetype Creation and Site Selection . . . 34
1.3.3 Simulation Outputs and Intervention Maps . . . 42
1.3.4 Site Selection Sensitivity Analysis . . . 44
1.4 Discussion . . . 46
1.5 Additional Plots . . . 52

2 Projecting Eradication Feasibility Using Archetypes 65
2.1 Introduction . . . 65
2.2 Methods . . . 67
2.2.1 Burden Change Due to Secular Trends . . . 69
2.2.2 Additional Burden Reduction Under Intervention Scenarios . . 71
2.2.3 Burden Estimation . . . 73
2.3 Results . . . 75
2.4 Discussion . . . 89
2.5 Additional Plots and Tables . . . 94

II Estimating Insecticide-Treated Net Coverage 107

3 Maps and Metrics of Insecticide-Treated Net Coverage in Africa: Access, Use, and Nets-Per-Capita, 2000-2020 111
3.1 Introduction . . . 111
3.2 Definitions . . . 114
3.3 Methods: Stock and Flow . . . 118
3.3.1 Background . . . 118
3.3.2 Data . . . 121

3.3.3 Model . . . 131
3.3.4 Sensitivity Analysis . . . 141
3.4 Methods: Spatiotemporal Regression . . . 142
3.4.1 Data . . . 143
3.4.2 Model . . . 144
3.4.3 Covariate Selection . . . 148
3.4.4 Uncertainty Propagation . . . 149
3.4.5 Access vs NPC . . . 151
3.4.6 Relative Gain . . . 151
3.5 Results . . . 151
3.5.1 Access and Use . . . 151
3.5.2 Access vs Nets-Per-Capita . . . 155
3.5.3 Net Retention Time . . . 159
3.5.4 Relative Gain: Optimising Access or Use? . . . 163
3.5.5 Stock and Flow Model Crop-to-Access Regression Coefficients 165
3.5.6 Stock and Flow Sensitivity Analysis . . . 165
3.5.7 Geospatial Regression Coefficients and Model Validation . . . 169
3.5.8 Example of Exceedance Surfaces for Uncertainty Visualization 169
3.6 Discussion . . . 173
3.7 Additional Plots and Tables . . . 180
3.7.1 Stock and Flow Model Fit by Country . . . 180
3.7.2 Stock and Flow Sensitivity Analysis by Country . . . 194
3.7.3 Sources for Report-Only Surveys . . . 203

4 Predicting the Impact of COVID-Related Service Disruptions on Insecticide-Treated Net Coverage, 2020 215

4.1 Introduction . . . 215

4.2 Methods 217
4.2.1 2020 Data 218
4.2.2 2020 Model 220
4.3 Results 221
4.4 Discussion 228

Conclusion **233**

Attributions and Publications **241**

Acknowledgements **245**

List of Abbreviations

AIS AIDS Indicator Survey

AL CM Artemether-Lumefantrine Case Management

ALMA African Leaders Malaria Alliance

ATSB Attractive Targeted Sugar Bait

BMGF Bill and Melinda Gates Foundation

CHAI Clinton Health Access Initiative

CHIRPS Climate Hazards Group InfraRed Precipitation with Station data

cITN Conventional Insecticide-Treated Net

CLARA Clustering Large Applications

COVID Coronavirus disease

DHS Demographic Health Survey

DP CM Dihydroartemisinin-Piperaquine Case Management

DRC Democratic Republic of the Congo

EA&P Anemia and Parasitemia Survey

GMM Gaussian Mixture Model

GMRF Gaussian Markov Random Field

HBHI High Burden High Impact

IDM Institute for Disease Modeling

IHME Institute for Health Metrics and Evaluation

INLA Integrated Nested Laplace Approximation

IQR Interquartile Range

IRS Indoor Residual Spraying

ITN Insecticide-Treated Net

LCME Lancet Commission for Malaria Eradication

LLIN Long-Lasting Insecticide-Treated Net

LOO Leave-One-Out

mAb Monoclonal Antibody

MAP Malaria Atlas Project

MARA Mapping Malaria Risk in Africa

MICS Multiple Indicator Cluster Survey

MIS Malaria Indicator Survey

MOP Malaria Operational Plan

NMCP National Malaria Control Program

NPC Nets-Per-Capita

PAR Population at Risk

PEV Pre-Erythrocytic Vaccine

$\boldsymbol{PfPR_{2\text{-}10}}$ *Plasmodium falciparum* parasite rate among 2-10 year olds

PIT Probability Integral Transform

PMI President's Malaria Initiative

RMSE Root Mean Squared Error

$\mathbf{SAG_{ME}}$ Strategic Advisory Group for Malaria Eradication

SSP/RCP Shared Socioeconomic Pathways/Representative Concentration Pathways

SVD Singular Value Decomposition

TBV Transmission-Blocking Vaccine

TSI Temperature Suitability Index

WAIC Watanabe-Akaike Information Criterion

WHO World Health Organization

List of Figures

1.1 Distribution of covariates for Singular Value Decomposition (SVD) across all pixels before and after rescaling. 35

1.2 Proportion of variance explained by the first five singular vectors after Singular Value Decomposition (SVD). 36

1.3 K-means elbow plot for k between 3 and 15. 36

1.4 Top and side view of pixel values reprojected onto the first three singular vectors, after k-means clustering with a k of ten. 37

1.5 Cluster maps for k of 3 to 14. 39

1.6 Maps and time series for the ten-site setting. 41

1.7 Example intervention impact results for four of the 152 interventions tested. 43

1.8 Sensitivity of sites to biting intensity for vector control vs non-vector-control interventions. 45

1.9 Sensitivity analysis sites. 47

1.10 Sensitivity analysis example intervention curves. 48

1.11 Maps and variable summaries for a cluster count of 3. 53

1.12 Maps and variable summaries for a cluster count of 4. 53

1.13 Maps and variable summaries for a cluster count of 5. 54

1.14 Maps and variable summaries for a cluster count of 6. 54

1.15 Maps and variable summaries for a cluster count of 7. 54

1.16 Maps and variable summaries for a cluster count of 8. 55

1.17 Maps and variable summaries for a cluster count of 9. 55

1.18 Maps and variable summaries for a cluster count of 10. 55

1.19 Maps and variable summaries for a cluster count of 11. 56

1.20 Maps and variable summaries for a cluster count of 12. 56

1.21 Maps and variable summaries for a cluster count of 13. 56

1.22 Maps and variable summaries for a cluster count of 14. 57

2.1 Distribution of covariates for Singular Value Decomposition (SVD) across all pixels in the analysis. 76

2.2 Proportion of variance explained by the first five singular vectors after Singular Value Decomposition (SVD). 78

2.3 K-means elbow plot for k between 3 and 15. 78

2.4 Top and side view of pixel values reprojected onto the first three singular vectors, after k-means clustering with a k of six. 79

2.5 Cluster maps for k of 3 to 14. 80

2.6 Maps and time series for the six-site setting. 81

2.7 Projected malaria prevalence due to secular trends alone, holding interventions at 2016 coverage levels. 83

2.8 Selected intervention package results in terms of incidence rates and total cases. 83

2.9 Example intervention impact results for four of the 152 interventions tested. 85

2.10 Sensitivity of sites to biting intensity for vector control vs non-vector-control interventions. 86

2.11 Intervention impact maps of $PfPR_{2\text{-}10}$ for the interventions presented to eradication project stakeholders. 88

2.12 Comparison of results using the archetypes and sites from Chapter 1 compared to Chapter 2. 89

2.13 Maps and variable summaries for a cluster count of 3. 94

2.14 Maps and variable summaries for a cluster count of 4. 95

2.15 Maps and variable summaries for a cluster count of 5. 95

2.16 Maps and variable summaries for a cluster count of 6. 95

2.17 Maps and variable summaries for a cluster count of 7. 96

2.18 Maps and variable summaries for a cluster count of 8. 96

2.19 Maps and variable summaries for a cluster count of 9. 96

2.20 Maps and variable summaries for a cluster count of 10. 97

2.21 Maps and variable summaries for a cluster count of 11. 97

2.22 Maps and variable summaries for a cluster count of 12. 97

2.23 Maps and variable summaries for a cluster count of 13. 98

2.24 Maps and variable summaries for a cluster count of 14. 98

3.1 Insecticide-Treated Net (ITN) model summaries. 115

3.2 Long-lasting insecticide-treated net distribution by type, including direct data sources and interpolated values. 124

3.3 Nationally-representative surveys included in this analysis, listed by country, year, and survey type. 126

3.4 Spatial stationarity of deviation metrics compared to original metrics. 146

3.5 National-level insecticide-treated net access and use time series. . . . 153

3.6 Maps of Insecticide-treated net access, use rate, and Nets-Per-Capita (NPC) in 2020, with associated uncertainty. 156

3.7 Insecticide-Treated Net (ITN) access deviation by country. 157

3.8 Nets-per-capita deviation by country. 157

3.9 Insecticide-Treated Net (ITN) use gap by country. 158

3.10 Insecticide-Treated Net (ITN) use rate by country. 158

3.11 Relationship between access and Nets-Per-Capita (NPC) in 2020. . . 160

3.12 Comparison of modeled and raw data relationship between nets-per-capita and Insecticide-Treated Net (ITN) access. 161

3.13 Insecticide-Treated Net (ITN) es optimal Access by country. 161

3.14 Long-Lasting Insecticide-Treated Net (LLIN) median retention times. 164

3.15 Magnitude of change in Insecticide-Treated Net (ITN) use possible from increasing use rate versus increasing access. 166

3.16 Relative gain example from 2015. 167

3.17 Out-of-sample root mean squared error evolution in sensitivity analysis. 170

3.18 Median retention time (or "half life") evolution in sensitivity analysis. 171

3.19 Geospatial regression coefficients. 172

3.20 Geospatial regression model fits. 172

3.21 Geospatial regression Probability Integral Transform (PIT) distributions. 173

3.22 Example 2020 exceedance surfaces. 174

3.23 Posterior density and parameter trace plots for Angola. 180

3.24 Posterior density and parameter trace plots for Benin. 181

3.25 Posterior density and parameter trace plots for Burkina Faso. 181

3.26 Posterior density and parameter trace plots for Burundi. 181

3.27 Posterior density and parameter trace plots for Cameroon. 182

3.28 Posterior density and parameter trace plots for Central African Republic. 182

3.29 Posterior density and parameter trace plots for Chad. 182

3.30 Posterior density and parameter trace plots for Comoros. 183

3.31 Posterior density and parameter trace plots for Congo (Republic of). 183

3.32 Posterior density and parameter trace plots for Cote d'Ivoire. 183

3.33 Posterior density and parameter trace plots for Democratic Republic of the Congo. 184

3.34 Posterior density and parameter trace plots for Djibouti. 184

3.35 Posterior density and parameter trace plots for Equatorial Guinea. . . 184

3.36 Posterior density and parameter trace plots for Eritrea. 185

3.37 Posterior density and parameter trace plots for Ethiopia. 185

3.38 Posterior density and parameter trace plots for Gabon. 185

3.39 Posterior density and parameter trace plots for Gambia. 186
3.40 Posterior density and parameter trace plots for Ghana. 186
3.41 Posterior density and parameter trace plots for Guinea. 186
3.42 Posterior density and parameter trace plots for Guinea-Bissau. 187
3.43 Posterior density and parameter trace plots for Kenya. 187
3.44 Posterior density and parameter trace plots for Liberia. 187
3.45 Posterior density and parameter trace plots for Madagascar. 188
3.46 Posterior density and parameter trace plots for Malawi. 188
3.47 Posterior density and parameter trace plots for Mali. 188
3.48 Posterior density and parameter trace plots for Mauritania. 189
3.49 Posterior density and parameter trace plots for Mozambique. 189
3.50 Posterior density and parameter trace plots for Niger. 189
3.51 Posterior density and parameter trace plots for Nigeria. 190
3.52 Posterior density and parameter trace plots for Rwanda. 190
3.53 Posterior density and parameter trace plots for Senegal. 190
3.54 Posterior density and parameter trace plots for Sierra Leone. 191
3.55 Posterior density and parameter trace plots for Somalia. 191
3.56 Posterior density and parameter trace plots for South Sudan. 191
3.57 Posterior density and parameter trace plots for Sudan. 192
3.58 Posterior density and parameter trace plots for Tanzania. 192
3.59 Posterior density and parameter trace plots for Togo. 192
3.60 Posterior density and parameter trace plots for Uganda. 193
3.61 Posterior density and parameter trace plots for Zambia. 193
3.62 Posterior density and parameter trace plots for Zimbabwe. 193
3.63 Sensitivity analysis results for Ghana. 195
3.64 Sensitivity analysis results for Malawi. 196

3.65 Sensitivity analysis results for Nigeria. 197

3.66 Sensitivity analysis results for Rwanda. 198

3.67 Sensitivity analysis results for Senegal. 199

3.68 Sensitivity analysis results for Sierra Leone. 200

3.69 Sensitivity analysis results for Tanzania. 201

3.70 Sensitivity analysis results for Zambia. 202

4.1 Time series comparison of Insecticide-Treated Net (ITN) use under different 2020 ITN distribution scenarios. 223

4.2 Comparison of national-level Insecticide-Treated Net (ITN) use with and without mass campaigns in 2020 for all countries with a mass campaign planned. 224

4.3 Maps of estimated Insecticide-Treated Net (ITN) use in 2020 and use as a proportion of the "Business as Usual" scenario. 225

4.4 Impact of Insecticide-Treated Net (ITN) and antimalarial service disruption scenarios on malaria incidence and mortality in 2020. 226

4.5 Impact of Insecticide-Treated Net (ITN) and antimalarial service disruption scenarios on malaria incidence and mortality in 2020. 227

List of Tables

1.1 EMOD mosquito species parameters. 31

1.2 EMOD intervention descriptions. 33

1.3 EMOD Site Descriptions. 47

2.1 Projected environmental covariate sources. 71

2.2 EMOD Site Descriptions. 74

2.3 Difference between old and new method. 75

2.4 Interventions and methods comparisons between Chapter 1 and Chapter 2. 103

3.1 Custom, country-specific strategies for interpolating missing net distribution data. 123

3.2 Surveys used in ITN models. 131

3.3 Stock and flow parameter definitions: indices and data. 133

3.4 Stock and flow parameter definitions: modeled values. 134

3.5 Parameter definitions for conversion of net crop to net access. 138

3.6 Countries included in sensitivity analysis. 141

3.7 Covariates used in all spatiotemporal models. 145

3.8 Covariates used in covariate selection tests. 150

3.9 Validation metrics for covariate sets. 150

3.10 Median retention times and 95% CI's for LLINs from the stock and flow model. 163

3.11 Regression coefficients for determining the proportion of households with no nets, ρ_h. 165

3.12 Regression coefficients for determining mean nets per household, μ_h. . 168

3.13 Sources for those surveys used in ITN models in which only report-level data were available. 207

4.2 Details of 2020 routine and mass ITN distribution numbers, by country. 220

4.1 Scenario names and descriptions for COVID mitigation analysis. . . . 221

4.3 Median LLIN retention times in this analysis that are meaningfully different from those in Chapter 3 due to differences in survey data availability. 222

Abstract

This thesis describes the development and refinement of two frameworks for usefully combining mechanistic and geospatial models in malaria research. The first framework, focused on malaria transmission, uses machine learning tools and spatial data to identify key features for a mechanistic transmission model, then utilizes spatial data again to place mechanistic model results in a global context. This strategy is novel, flexible, customizable, and computationally efficient. The second, focused on insecticide-treated net coverage, uses a mechanistic commodity tracking model as the backbone for a spatiotemporal regression model in order to appropriately capture different countries' unique histories of net procurement, distribution, and retention. Examples of the on-the-ground policy impact of this work are included for stakeholders such as the World Health Organization, the Clinton Health Access Initiative, and the Bill and Melinda Gates Foundation. Both of these modeling frameworks showcase the value of combining the flexibility and versatility of mechanistic modeling with the statistical rigor and geographical tools of geospatial modeling.

Introduction

> I'm just trying to find a slightly useful way of saying there are things I don't understand.
>
> Marilynne Robinson
> *Gilead*

Models, in the sense of simplified descriptions of systems that help us understand and move through the world, are foundational to human experience. We each carry inside us nesting and intertwined models of the world which we tap into constantly, whether navigating through our city's traffic patterns, performing the chemistry of meal preparation, following the social norms of our homes, or exploring the virtual spaces accessible through a screen. One of the most common types of model is the map, which simplifies space and location and encodes information purposively to serve a specific goal or set of goals. All of our internal models, spatial and otherwise, are constantly changing and updating in response to new information. One needs only to misremember the number of steps outside one's front door to be reminded of how jarring it can be to make decisions based on a poorly-representative model.

Formalized models, which use mathematics or statistics to explicitly map out theorized relationships, have a wide range of uses. First, formal model descriptions can be used for communication and comparison to assess how different people conceptualize the same system. A model can be used for hypothesis testing in both directions– when a model and data do not line up, both the model and the data

may deserve careful reanalysis and consideration. A model that has been shown to effectively capture the phenomena it claims to represent can be used to predict what effect different decisions might have. A model that represents a past event can be used to explore counterfactual scenarios of what might have been. The process of creating and calibrating a model can reveal important information about which features of the world are salient and which are incidental with respect to a phenomenon of interest.

Public health and epidemiology have a close historical relationship with modeling. The field of statistics as it is currently recognized was developed extensively in the 17th century to assess demographic data related to births and mortality, linking these statistical practices closely to questions of population health [1]. In epidemiology, mapping was a crucial tool in understanding and preventing outbreaks years before the germ theory of disease elucidated the causal pathways of pathogen transmission [1, 2]. The use of modeling and mapping in public and global health has come in and out of vogue in the centuries since, but recent decades have seen an ever-increasing push towards data collection, analysis, and modeling as a technological solution to many health-related challenges worldwide.

As a disease, malaria far predates any of these statistical or mathematical developments, having co-evolved with mosquitoes, reptiles, birds, and mammals over millennia. In the last 150 years, however, the histories of malaria and mathematical modeling have been deeply intertwined. Sir Ronald Ross (1857-1932), an Indian-born British doctor, is credited not only for discovering that malaria is a parasite transmitted by mosquitoes, but also for laying the foundations of disease transmission modeling as a theory [1] and for creating the first mathematical descriptions of malaria transmission [3]. Ross' ties to India serve as a reminder of the deeply-ingrained relationship between colonialism, disease, and Western-led quantitative research that still thrives today [4, 5]. The work of Ross, Anderson Gray McK-

endrick, William Ogilvy Kermack, George Macdonald, and others [1] were some of the first formal examples of what are now called mechanistic disease transmission models: mathematical constructs that explicitly replicate the pathways of disease transmission as described by transmission equations and actuated through rates or probabilities. Mechanistic modeling has thrived with the advent of scientific computing, which has allowed the actualization and calibration of such models via simulation. Mechanistic models are now extensively used to study and intervene on dozens of diseases, not just malaria [6].

The computational revolution also drove the expansion and sophistication of spatial statistical methodologies both in and outside of public health [7]. In recent decades, malaria once again became the poster child for a novel mathematical public health technique when the Mapping Malaria Risk in Africa group and Malaria Atlas Project (MAP) began using statistical models that combined cross-sectional survey data of malaria with remotely-sensed climate and land use data to publish high-resolution maps of estimated malaria transmission intensity in Africa and globally. These products became immensely popular, and have inspired the creation of high-resolution maps of many other diseases and conditions including HIV [8], child mortality [9], snake bites [10], and more.

Mechanistic and geospatial modeling methods are both widely used in the malaria field today, and the overlap between the two is far from negligible. Mechanistic models often have a spatial component whose parameters are informed by high-resolution geospatial surfaces [11, 12], while conversely geospatial modelers have used the relationships learned from mechanistic models to inform prevalence-to-incidence conversion calculations [13] and more. Practical and administrative technicalities, however, often hamper the impact of these interdisciplinary projects. As one example, scientists working outside of MAP may only have access to publicly-available spatial estimates of any given metric from the group, which may in some cases be several

years out of date. As another, while the Institute for Disease Modeling (IDM)'s mechanistic malaria model EMOD is publicly available, at present only institute employees and a restricted list of external users have access to the software support and cloud infrastructure that allows for large-scale use of the model.

The four years of work represented by this thesis took advantage of a rare opportunity for a single scientist to be fully affiliated with both the world's leading spatial malaria modeling group at MAP and a top-tier mechanistic modeling organization at IDM. Because both of these organizations are focused on working with policy stakeholders to answer questions that assist decision making, the goal of this doctoral work was explicitly focused on developing, testing, and applying integrated mechanistic/spatial modeling frameworks for the purpose of providing policy advice with broad-reaching impact.

Specifically, this document presents two major analytical undertakings that showcase the strength of merging mechanistic and geospatial models into a single framework. Both methods focus on *Plasmodium falciparum* malaria in Africa, are continental in scope, and have had a demonstrable impact on international malaria policy. The first, in Part I, focuses on mechanistic disease transmission models and transmission intensity maps, and showcases an innovative framework for assessing continental-scale intervention impact. The second, in Part II, centers a mechanistic model of commodity supply chains, the output of which is embedded within a geospatial regression infrastructure to create high-resolution time series of insecticide-treated net coverage in sub-Saharan Africa. These two analyses are entirely original or major adaptations of existing work, respectively, and address a myriad of challenges: the theoretical difficulties of combining disparate modeling philosophies, the technical complications of high-resolution spatial modeling within the limits of today's computational infrastructures, and the linguistic and graphical nuances of effectively communicating modeled results to stakeholders from a range

of different backgrounds. These analyses were made possible thanks to my insider access within both MAP and IDM, but over the course of this work my colleagues and I have also prioritized making such collaborations more accessible to others. This is evidenced in Part I by the ability of the archetypes framework to be used with any combination of spatial estimates and mechanistic models, and in Part II by extensive documentation and public availability of both results and code.

Chapter 1 introduces a novel archetypes-based approach which harnesses the strength of both mechanistic and geospatial models to estimate continent-level burden impacts of different malaria interventions. It first places the importance of this work in historical context, then describes a theoretical framework and an example model configuration. Use cases, best practices, and limitations are discussed.

Chapter 2 demonstrates a case study of the archetypes framework applied to the question of malaria eradication feasibility within the next 30 years. The results of this analysis played a central role in the first international effort to set a target date for malaria eradication since the mid 20th century. A brief history of malaria eradication efforts is described, followed by a detailed description of the modeling analysis and a reflection on the important yet complicated relationships between modelers and policymakers, especially for time-sensitive projects.

Chapter 3 shifts gears away from transmission modeling and toward the modeling of insecticide-treated net coverage. A framework which combines a mechanistic supply chain model with a geospatial net coverage model is introduced and placed in historical context. This model explicitly differentiates between net access and net use. The model is described in detail, and various maps and metrics of net use and their implications for policymakers is discussed.

Chapter 4 describes an emergency use case of the insecticide-treated net modeling framework to assess coverage changes in the face of COVID-related service

disruptions. This work was commissioned by the World Health Organization and the Bill and Melinda Gates Foundation in spring of 2020 and contributed to international malaria control policy decisions for the remainder of the year. The potential impact on burden of both net and antimalarial drug-related service disruptions are described and discussed.

The concluding chapter reviews the impact of this work, focusing on both policy decisions and stakeholder interactions that have resulted from these analyses. This chapter also highlights future analyses and discusses some moral responsibilities of researchers working within inequitable global power structures. Reviews of the relevant literature are included at the beginning of each chapter.

Bibliography

1. Brauer, F. Mathematical epidemiology: Past, present, and future. *Infectious Disease Modelling* **2,** 113–127. ISSN: 2468-0427 (2017).
2. Koch, T. *Disease maps : epidemics on the ground* 330. ISBN: 9780226449357 (University of Chicago Press, 2011).
3. Smith, D. L. *et al.* Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. *PLoS Pathogens* **8.** ISSN: 15537366 (2012).
4. Kim, H. The implicit ideological function of the global health field and its role in maintaining relations of power. *BMJ Global Health* **6.** https://gh.bmj.com/content/6/4/e005620 (2021).
5. Athni, T. S. *et al.* The influence of vector-borne disease on human history: socio-ecological mechanisms. *Ecology Letters.* ISSN: 14610248. https://onlinelibrary.wiley.com/doi/epdf/10.1111/ele.13675 (2021).
6. Lessler, J. & Cummings, D. A. T. Mechanistic Models of Infectious Disease and Their Impact on Public Health. *American Journal of Epidemiology* **183,** 415–422. ISSN: 0002-9262. https://academic.oup.com/aje/article/183/5/415/2197222 (2016).
7. Elliott, P. & Wartenberg, D. Spatial epidemiology: Current approaches and future challenges. *Environmental Health Perspectives* **112,** 998–1006. https://ehp.niehs.nih.gov/doi/10.1289/ehp.6735 (2004).
8. Dwyer-Lindgren, L. *et al.* Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017. *Nature* **570,** 189–193. ISSN: 1476-4687. https://www.nature.com/articles/s41586-019-1200-9 (2019).
9. Burstein, R. *et al.* Mapping 123 million neonatal, infant and child deaths between 2000 and 2017. *Nature* **574,** 353–358. ISSN: 1476-4687. https://www.nature.com/articles/s41586-019-1545-0 (2019).
10. Longbottom, J. *et al.* Vulnerability to snakebite envenoming: a global mapping of hotspots. *The Lancet* **392,** 673–684. ISSN: 0140-6736. http://www.thelancet.com/article/S0140673618312248/fulltext (2018).
11. Winskill, P., Walker, P. G., Griffin, J. T. & Ghani, A. C. Modelling the cost-effectiveness of introducing the RTS,S malaria vaccine relative to scaling up other malaria interventions in sub-Saharan Africa. *BMJ Global Health* **2,** 90. https://gh.bmj.com/content/2/1/e000090 (2017).

12. WHO. *World Malaria Report 2020* tech. rep. December (2020), 238.

13. Cameron, E. *et al.* Defining the relationship between infection prevalence and clinical incidence of Plasmodium falciparum malaria. *Nature Communications* **6,** 8170. ISSN: 2041-1723. http://www.nature.com/doifinder/10.1038/ncomms9170 (2015).

Part I

Assessing Intervention Impact

Chapter 1

An Archetypes Approach to Intervention Modeling

> They say there is nothing more difficult than to define a spiral in words; it's necessary, they say, to describe it in the air, with one's illiterate hands, using gestures spiraling slowly upwards to show how that abstract form, particular to coiled springs and certain staircases, appears to the eye.
>
> Fernando Pessoa
> *The Book of Disquiet*

1.1 Introduction

Malaria is one of humanity's oldest and most insidious ailments, co-evolving with mosquitoes and humans over millions of years [1]. This long history, combined with its complex transmission pathways, makes malaria a uniquely heterogeneous and environmentally sensitive disease. Understanding malaria in a given location requires not just an understanding of human behavior, movement, and demographics, but also a detailed knowledge of resident mosquito species and their behavior, local

climate and hydrology, and the seasonal patterns of the landscape. This complexity, combined with its high burden throughout history, made malaria a natural subject for one of the earliest and most detailed mathematical disease models. In the 1950s, as the international health community was embarking with difficulty on the world's first disease eradication effort [2], mathematician and entomologist George Macdonald refined and popularized a tool first architected by Ronald Ross in the early 20th century stipulating a series of equations governing the fundamental dynamics of malaria transmission. The hope was that, by disentangling the different drivers of the disease and understanding the impact of each variable on transmission, modelers could assist policymakers in understanding the most effective interventions to target in order to achieve elimination and eradication [3].

This first global effort at malaria eradication did not succeed, but it laid the groundwork for a close relationship between malaria modeling and policy decisions that continues to this day. Ross and Macdonald's pen and paper equations have evolved into detailed computational models capable of simulating individual humans, mosquitoes, household and community structures, climates, immune responses, malaria interventions, and much more. The fundamental goal of these models, however, remains unchanged: to help decision-makers craft a malaria strategy that aligns best with the epidemiological, economic, and cultural circumstances at hand. Over the last decade of malaria control and elimination efforts, modeling has extensively supported decision-making at the local, national, and global scale [4–8].

In parallel to these simulation-based, or mechanistic, transmission models, a second class of mathematical model has proven invaluable for malaria policy planning. This is the spatiotemporal regression approach, a child of classic statistical regression methodologies. Like other regression methods, this strategy takes advantage of correlations between outcomes of interest (e.g. malaria prevalence) and other

covariates (often climate, vegetation, entomology, and demographics) to infer relationships and estimate outcomes of interest in spaces without observed data points. When the available data include detailed information of both the location and time of observation, spatiotemporal methods can use covariance structures to capture these spatial and temporal relationships for improved inference and prediction [9, 10]. The development of remote sensing methods, and subsequent public availability of dozens of different climatic and demographic variables at high resolution, helped push malaria to the cutting edge of disease transmission mapping. In 2008, building on the work of the Mapping Malaria Risk in Africa (MARA) project [11–14], the Malaria Atlas Project (MAP) was launched. MAP was the first group to utilize high-resolution satellite data alongside nationally-representative household surveys to generate high-resolution maps estimating malaria burden across Africa [15]. In the years since, MAP's work has expanded to cover the global scale [15], to estimate the burden of *Plasmodium vivax* [16, 17] as well as *falciparum* [18], and to estimate a wide range of other malaria-related metrics such as mosquito species distribution [19, 20], malaria intervention coverage [21–24], and accessibility of healthcare resources [25]. These maps are utilized by country programs, international policymakers, and malaria researchers to understand both the current state and the history of malaria in the world, and as a baseline for future projections.

In many ways, mechanistic and spatiotemporal models serve complementary purposes in malaria. Spatiotemporal models elucidate the past and present of disease burden and related metrics, which mechanistic models can then use for insight into the future. Maps provide a rich descriptive landscape, while mechanistic models contribute causal structure and exploration of counterfactual or hypothetical scenarios. As both modeling methods have grown in popularity and demand, there is increased interest in products that combine the two approaches, allowing for a spatially and temporally detailed exploration of the consequences of different policy decisions. This chapter introduces a novel methodology which combines the rich-

ness of spatiotemporal mapping with the rigor of mechanistic modeling to create a multi-purpose infrastructure for answering a range of questions in the malaria policy space.

Mechanistic models draw strength from their ability to replicate detailed relationships, a particularly crucial quality when exploring a disease as nuanced as malaria. However, this strength is also a limitation for two reasons. First, these detailed models require a large amount of input data to configure and calibrate to a particular setting. For example, for a modeling collaboration between ISGlobal in southern Mozambique and the Institute for Disease Modeling (IDM), a full calibration of the model utilized geolocated demographic and household surveys, facility-level data on clinical malaria, painstakingly-collected entomological data, and a travel survey to assess human movement in the region, to name just a few [Suresh et al., in preparation]. When tasked with simulating settings without such a rich data space, mechanistic models can be a challenge to appropriately configure. Second, mechanistic model complexity carries with it an associated computational cost. Depending on the population size and intra-model interactions with which a mechanistic model is configured, a single simulation can take minutes or hours to run, even on high-performance computing systems. As the number of locations and scenarios increases, it can quickly become prohibitive to run a full suite of simulations.

In contrast to these challenges, spatiotemporal methods are designed with incomplete data as an assumption, and perform well at high resolutions. These models explicitly take advantage of relationships between available variables to generate estimates for data-sparse areas, and statisticians in the field have developed advanced methods for efficient computation at high spatial and temporal resolution, especially on high-performance clusters [26, 27]. These statistical methods, however, crucially lack the explicit causal relationships that allow mechanistic models to so effectively test the consequences of different policies.

In the past, mechanistic modeling analyses working at the continental scale in Africa have utilized a range of methods to address these practical and computational challenges. Griffin et al. [28] modeled a range of stylized seasonality patterns and mosquito species mixes to assess different intervention strategies. While this approach showcases hypothetical scenarios, there is no explicit match between any true location and any stylized seasonal profile, making geographic trends or burden estimation impossible. Walker et al. [4] grouped rainfall, entomological, and transmission intensity spatial covariates into three, four, and 18 clusters respectively, and ran simulations on all possible combinations of these groups before re-projecting onto the pixel level. This exhaustive strategy does not take advantage of the spatial relationships between different covariates, which could reduce computational burden and generate more informative cluster properties.

The methodology presented in this chapter leverages the strength of both mechanistic and spatiotemporal methods to allow computationally-feasible generation of high-resolution maps that reflect mechanistically modeled scenarios. This process generates a range of archetypal seasonal and entomological profiles that can be useful in and of themselves for exploratory data analysis, while also generating an explicit link between these profiles and any given location across Africa. This mapping capacity provides a computationally efficient pathway for presenting results geospatially and in terms of expected burden change.

In this methodology, high-resolution, high-dimensional spatial covariates and machine learning methods are harnessed to generate a small number of spatially-explicit "archetypes" of malaria transmission, characterized by their covariate similarity. Next, mechanistic models are run on a representative site from each archetype, rather than on every pixel individually. Finally, spatial data and model results create a lookup table through which maps of intervention impact are generated.

Sorting many observations of high-dimensional data into groups by similarity is

a common problem in machine learning. A common solution is to deploy a two-step process of first reducing the dimensionality of the dataset, and then clustering this reduced-dimensional data into groups. Examples of this approach can be seen in oceanography [29], facial recognition [30], and sensor placement [31]. This approach is appealing for its flexibility, but requires a number of specific decisions that can materially impact results. Namely:

1. Which covariates should be selected, and how should they be standardized?
2. Which algorithm should be used for dimensionality reduction?
3. Which algorithm should be used for clustering, and how should the number of clusters be determined?

When clustered results are being used as inputs for mechanistic models and subsequent generation of new maps, additional questions emerge:

4. How should representative sites be selected from each cluster?
5. Which mechanistic model should be used to assess intervention impact?
6. How should results from representative sites be reprojected back onto other members of the group?

This chapter will review each of these questions in turn, beginning with a discussion of the choices available in each space and their implications. As a demonstration, it will describe an example configuration of the archetypes framework. This example was created as a post-hoc validation for an eradication feasibility exercise undertaken by the IDM and the MAP in 2018. The original work was conducted by request of international stakeholders curious whether eradication might be possible under highly optimistic environmental circumstances and malaria control scenarios.

It harnessed spatiotemporal climate and mosquito covariates, IDM's EMOD malaria transmission model, and MAP's malaria prevalence maps to generate estimates of malaria intervention impact at the 5km-by-5km pixel level for all of sub-Saharan Africa. As described in Chapter 2, the original eradication feasibility project utilized an earlier version of the framework, and the version presented here in Chapter 1 represents a validation check to ensure that similar results arise from a more in-depth approach. A sensitivity analysis exploring the representativeness of a given "representative site" is also described.

Throughout this document, the terms "archetypes strategy" and "archetypes framework" will be used to describe the general methodology of using dimensionality reduction and clustering to locate a subset of sites appropriate for mechanistic modeling. The term "example configuration" will be used to describe the specific set of choices within the broader strategy that are shown here for demonstration. The archetypes strategy is flexible and generalizable to any covariate set, mechanistic modeling platform, and geographic scale. The methods sections are subdivided into a general discussion and an explicit description of the example configuration parameters. The results section focuses on describing example configuration outcomes. Finally, the discussion section highlights lessons learned and other use cases of this archetypes framework.

1.2 Methods

In the sections below, each of the six methodological questions described in the introduction is explored in detail. General theory is described first, followed by a description of the specific choices made in the example configuration. For the example configuration, simplicity was favored, but more complex options are noted.

1.2.1 Covariate Selection

General

The covariates used for clustering should be selected mindfully to capture the types of variation to which the transmission model is most sensitive. For malaria, these should almost always include covariates that capture the different seasonality patterns of different malaria-endemic regions, as this directly impacts disease seasonality. For models that explicitly simulate different mosquito behaviors by species, covariates describing relative species abundance are also valuable. Baseline malaria transmission intensity is another useful source of input data, but its inclusion depends on the use case. The example configuration described below intentionally excludes it from the clustering process, instead running simulations for each archetype over a range of transmission intensities. Other possible covariates of interest include intervention history, health care accessibility, population demographics, and other behavioral inputs. Covariate data must geographically cover the region of interest, and should be utilized on the same spatial resolution as the final results. Covariate data for malaria will often also include a temporal component showing either seasonality or secular time.

Covariates are collected into a "stack" of spatial or spatiotemporal input files measuring different metrics on different scales. It is important to normalize or rescale these covariates prior to performing dimensionality reduction to avoid an artificial effect due to differently-scaled inputs. Even after normalization, however, a decision must be made regarding the relative weight of different inputs. For instance, in the example configuration below, clustering covariates included 12 precipitation layers, 12 temperature layers, and three mosquito species relative abundance layers. These covariates were not differentially weighted prior to dimensionality reduction, meaning that the rainfall and temperature layers had a stronger impact on results

than the mosquito abundance layers. This choice was acceptable for the work at hand, but different circumstances might encourage different weighting.

Example Configuration

Environmental covariate data on precipitation and air temperature were sourced from the ERA5 project (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5), a global reanalysis that generates internally-consistent estimates for a wide range of climate parameters across the globe. Specifically, monthly mean total precipitation and 2-meter air temperature from 2000 to 2018 were downloaded. Monthly means were averaged across the time series to generate synoptic 12-month trends. Because ERA5 spatial resolution is 0.25°-by-0.25°(approximately 30km-by-30km at the equator), covariates were resampled down to the 5km-by-5km level and realigned to match the standard spatial specifications used by MAP.

Covariate data on the relative abundance of *Anopheles arabiensis*, *funestus*, and *gambiae* were obtained from Sinka et al. [20]. These values are static estimates and did not require resampling as they already met MAP spatial specifications.

All covariate rasters were masked to align with MAP estimates of *Plasmodium falciparum* transmission limits, and pixel-level values extracted into a nonspatial dataset for further analysis. Once extracted, pixel-level values of air temperature and precipitation were rescaled to fall between zero and one. Because they contained some extreme values that severely skewed the distribution, the most extreme 1% of precipitation values were reassigned to the 99th percentile value before being rescaled. Relative vector abundance values are proportions, and therefore these values already fell between zero and one and did not require rescaling. See Figure 1.1 for covariate distributions before and after rescaling.

1.2.2 Dimensionality Reduction

General

If the covariate selection process only locates a few variables of interest, a dimensionality reduction step may not be necessary. However, beyond the point that covariates can comfortably be plotted together (four or five layers at most), dimensionality reduction can be a valuable tool both for data exploration and for more effective clustering. The goal of these techniques is to collapse high-dimensional data into a lower-dimensional space in a way that preserves as much of the original data variation as possible, creating a denser and richer database from a sparser one. A wide range of dimensionality reduction techniques are available in machine learning software packages.

This analysis utilized Singular Value Decomposition (SVD), a strategy similar to Principal Components Analysis which locates the set of orthogonal vectors in a high-dimensional dataset that cover the most variance in the data, and returns those vectors in order of variance explained [32, 33]. SVD is evaluated by plotting how much variance in the original dataset is explained by the most informative singular vectors. If a small number of vectors covers a large portion of the initial variance, then only that smaller-dimensional dataset need be retained for further analysis. While straightforward to implement, SVD is not a time-series method, meaning that dimensionality reduction does not explicitly take into account causal correlations between the different monthly layers of climate data. Other strategies, such as Fourier transforms or dynamic mode decomposition, are worth consideration if simple methods such as SVD do not yield informative results, or if the input data cover long time series rather than synoptic trends.

Example Configuration

The final covariate dataset was comprised of 27 dimensions: 12 layers each of precipitation and air temperature data and three layers of relative vector abundance data. SVD was applied using the `svd` command in `R` version 3.6.2. Points that are close together in this reprojected space are expected to be similar in terms of their precipitation, temperature, and relative vector proportions. Upon running SVD, the first three singular vectors explained over 95% of the variation in the full dataset, and were retained for clustering (Figure 1.2).

1.2.3 Clustering

General

The literature and theory on clustering algorithms is vast [34, 35], including many excellent tutorials for beginners. Briefly, clustering strategies are unsupervised learning methods that group data points together based on proximity in all provided dimensions.

The example configuration utilized k-means for clustering. This algorithm benefits from simplicity and relatively quick runtime using an iterative geometric search, but suffers two major setbacks. First, the choice of cluster count k must be specified by the user, which can lead to nonintuitive cluster groupings if an inappropriate k is selected. Intuitively, the goal of k-means is to minimize data variance within clusters while maximizing it between clusters. A heuristic but common solution for addressing the cluster count problem is the use of "elbow plots", in which cluster count is plotted against the ratio of between-cluster variance and total variance. This ratio is monotonically increasing, since the cluster count that maximizes this ratio is a k equal to the number of data points, but the goal is to find a k for which the plot

makes an "elbow" and begins to increase more slowly. Other strategies for selecting cluster counts for k-means include the use of Jaccard, Dunn, and Silhouette indices [35].

A second weakness of k-means is its radial search methodology. Because this clustering algorithm minimizes the Euclidean distance between a cluster centroid and the data points around it, clusters that are tightly grouped in all dimensions will be captured more effectively than elliptical or other cluster shapes. Heuristically, this did not appear to affect the algorithm's ability to find distinct groups, but nonspherical methods such as Clustering Large Applications (CLARA) [36] or Gaussian Mixture Model (GMM)s may be more appropriate in some settings. GMMs, which utilize a Bayesian fitting procedure, have the added benefit of allowing for the calculation of information criteria, giving a more principled and less heuristic way to select cluster counts compared to k-means.

Example Configuration

Having reduced the dimensionality of the dataset from 27 to three, the next step was to classify points into transmission archetypes representing unique environmental and entomological patterns of relevance to malaria transmission. To identify these distinct archetypes, k-means clustering was performed on the reduced-dimensional dataset output by SVD. K-means was run using the MacQueen algorithm [37] on the reduced-dimensional dataset over a range of possible cluster counts from 3 to 15, and elbow plots computed. Unfortunately, these plots did not show a strong elbow (Figure 1.3), so a cluster count of 10 was selected as a compromise between relatively high variance ratio and a visually intuitive and distinguishable set of archetypes (Figure 1.4, Figure 1.5, Figure 1.6). K-means was applied using the `kmeans` command in the `stats` R package version 4.0.5. See the Results section for a more in-depth description of k-means outputs.

1.2.4 Selecting Representative Sites

General

The clustering step groups the data into archetypes. Next, a set of specific covariate values must be selected for each archetype to use as inputs into the transmission model. One viable option would be to take the mean or median across all data points and use these as model inputs for the archetype, acknowledging that this measure of center will not reflect any true location and may be sensitive to outliers in the clustering process. Using the values at each cluster centroid would be less sensitive to outliers, but still not reflect any true physical location. A third alternative, used in the example configuration, is to select the data point closest to each k-means cluster centroid and use this as a representative site for the archetype as a whole. Visual examination of the k-means clusters (Figure 1.4) shows that the third approach is nearly identical to the second in the example configuration, as there is generally a trivial distance between the true cluster centroids and their nearest data point. Sensitivity analysis was conducted to assess the acceptability of this approach.

Example Configuration

The data point closest to each cluster centroid (determined using k-Nearest Neighbors, with k set to one) was selected as the site whose input data would be used for simulation. K-nearest neighbors was applied using the `get.knn` command in the `FNN` R package version 1.1.3.

Site Selection Sensitivity Analysis

In the main analysis, a single centrally-located point is used as a proxy for all other locations in that cluster. A sensitivity analysis was conducted to assess how other

locations in each cluster behaved under the same intervention scenarios. Ten pixels were randomly chosen from each cluster, and all intervention scenarios were re-run on these 1,000 points in the same way that they were on the original ten. Results were compared to those from the cluster centroid sites alone to assess within-cluster variance.

1.2.5 Selecting and Configuring a Malaria Model

General

A number of mechanistic malaria models are commonly used by groups consulting with national or international malaria policy groups. These include OpenMalaria [38], EMOD [7], the Imperial model [28], and VCOM [39]. These models vary with respect to their structure, ease of use, and public accessibility and can produce measurably different estimates under similar initial conditions [40]. Modelers should have a close familiarity with their model of choice prior to embarking on the exercises described in this document. In particular, it is necessary to understand the input parameters to which each model is most sensitive and how this sensitivity impacts covariate and archetype selection; how to appropriately parameterize each site's demographics, climate, entomology, and interventions; and how to run checks to diagnose unexpected model behavior before acting upon any conclusions from model results.

When pursuing the archetypes strategy, modelers will utilize values from the clustering covariates for some model inputs, but must choose how to initialize all variables not included in the clustering covariates. For instance, in the example configuration, the clustering analysis provided input precipitation, temperature, and entomology variables, but not demographic, intervention history, or transmission intensity variables. As described in detail below, this analysis held most other vari-

ables static across all sites, and ran simulations at a range of initial transmission intensities, but different use cases might call for different input parameters in different representative sites.

Example Configuration

This modeling analysis was conducted using the EMOD simulation software produced by IDM. See Section 1.2.7 for an introduction to EMOD. All simulations were run with an initial population of 2,000 to balance population size with computational constraints, a birth rate of 36.3 live births per 1,000 people per year from the World Bank Database, and a complementary mortality rate to ensure a stable population over time. Each human in the model was assigned a unique risk of being bitten by a mosquito, such that the distribution of biting risk in the community overall was exponential [41].

Simulations were initialized by running one 39-year intervention-free simulation for each of 50 initial transmission intensities and 10 random seeds, generating a collection of 500 baseline populations upon which to test intervention impact. Transmission intensities were varied by scaling mosquito larval habitat capacity, which linearly impacts the number of adult mosquitoes and the level of malaria transmission in the absence of interventions. The intervention-free immunity establishment period should approximate the length of a human life, and is frequently set to 50 years. A 39-year period was selected because this is the duration of time for which climate data were available from ERA5. Ten is a common choice for number of random seeds to test in EMOD, and has been shown to cover the variation in most parameters well. Fifty transmission intensities were chosen to thoroughly cover variability in this important parameter. One hundred fifty-two different intervention packages were tested, covering a wide range of vector control, drug, and vaccine-based malaria prevention strategies. These intervention parameters were originally

part of a project to test eradication feasibility in best-case scenarios, and therefore often represent coverage levels or policies more rigorous than those commonly in use today (for example mass bednet campaigns every year instead of every three years). Each intervention was run for three years, and its efficacy assessed by mean *Plasmodium falciparum* parasite rate among 2-10 year olds ($PfPR_{2\text{-}10}$) in the final year of the intervention compared to the final year of the intervention-free simulation.

Climate inputs of air temperature, precipitation, and relative humidity were constructed from the publicly-available ERA5 climate reanalysis model. For each representative site, daily climate data from 1980 to 2018 were downloaded from ERA5. The "2m temperature" channel was used for air temperature, the "total precipitation" channel for precipitation, and the "2m temperature" and "2m dewpoint_temperature" channels used to calculate relative humidity according to a method developed by the company Vaisala [42]. The intervention-free simulations to establish population immunity were run using climate data from 1980 to 2018, and the three-year intervention simulations were run using climate data from 2016 to 2018.

Mosquito parameters can be found in Table 1.1, and intervention descriptions and parameters in Table 1.2 and Table 2.4.

1.2.6 Recreating Maps

General

Once model simulations have run for each archetype's representative site, these individual-site results must be reprojected back onto pixel level results. The methodology described below constructs a lookup table to convert from baseline per-pixel transmission to intervention-impacted transmission in each archetype. If transmis-

sion intensity was included as a clustering covariate, the lookup table approach would differ in its details but be similar in essence.

Example Configuration

Once all EMOD simulations were complete, maps of intervention coverage were reconstructed as follows. First, a pixel-level map of $PfPR_{2\text{-}10}$ was selected from the MAP library. Because these EMOD simulations were initiated from intervention-free scenarios, the MAP estimate of $PfPR_{2\text{-}10}$ in 2000 was selected as a proxy for malaria prevalence in the absence of any interventions. Then, for each intervention scenario and each pixel p in the selected map, the following steps were taken:

1. The archetype a to which the pixel p belongs was identified;
2. A spline was generated between data points of initial and final $PfPR_{2\text{-}10}$ across all transmission intensities in archetype a;
3. This spline was used to map the initial $PfPR_{2\text{-}10}$ corresponding to that of pixel p and archetype a to the final $PfPR_{2\text{-}10}$ in that intervention scenario;
4. This final $PfPR_{2\text{-}10}$ was logged as the "intervention impact" of pixel p.

This allows for the reconstruction of maps hypothesizing the potential impact of different interventions across the continent.

1.2.7 Details on the EMOD Software

EMOD: An Overview

EMOD (https://idmod.org/documentation, [43]) is an individual-based stochastic mechanistic modeling software developed by IDM. While EMOD is capable of

modeling a number of infectious diseases, only the malaria framework is described in detail here.

Operating on a daily time step, EMOD simultaneously tracks multiple simulation layers, including but not limited to: mosquito habitat and life cycle, mosquito behavior and movement, human population demographics, human movement, human immune responses to malaria, and malaria interventions that disrupt any of the aforementioned layers. Typically, humans are modeled as individuals while mosquitoes are modeled as cohorts, though there are options to model mosquitoes individually for extremely small-scale simulations. As an individual-based model, most input parameters are defined via probability distributions which are sampled at the appropriate points. Spatially distinct communities can be defined, with or without movement between them. While some heterogeneity in risk and behavior can be imposed within communities, under most circumstances a community behaves as a well-mixed population.

Mosquito Habitat, Life Cycle, Behavior, and Movement In EMOD, species-specific mosquito population size is driven by the size of larval habitat available. In the absence of interventions, mosquito abundance in turn drives community-level malaria prevalence. Larval habitat size over time can be specified manually, if entomological information is available to inform this estimate, or can be driven by climate files indicating daily air temperature, rainfall, and relative humidity in the site being modeled. If the climate-driven option is selected, a decision must also be made about the types of larval habitat available in the model, allowing for hydrology-based conversions between climate and mosquito emergence and mortality. For more detail see Eckhoff [7]. Species-specific anthropophily and endophily values can also be specified. In the cohort-based model, mosquitoes do not travel from one community to another. See Table 1.1 for a list of vector parameters utilized in the example

configuration.

Name	Anthropophily (%)	Endophily (%)
arabiensis	65	50
funestus	65	85
gambiae	85	85

Table 1.1: EMOD mosquito species parameters.

Human Population Demographics Humans are modeled individually in EMOD, with each person assigned an age, sex, and "home" community at model initialization. Age distributions and birth and death rates can be specified at model initiation. In addition, individuals can be assigned custom properties related to disease risk or exposure to better capture heterogeneity within communities. For example, in the example configuration, 10% of individuals were assigned to a custom "No Nets" category, to indicate that they would always be skipped during insecticide-treated bednet campaigns, and each individual was assigned a unique biting risk such that the distribution of biting risk in the community overall was exponential [41]. While various types of human movement between communities can be specified in the model, this analysis did not allow for human movement to keep the effects of any one archetypal site clear.

Human Immune Responses to Malaria and Accounting for Malaria Population Immunity Every time a person is infected with malaria in EMOD, they trigger a full intrahost infection cascade. Among the variables tracked are parasitemia, gametocytemia, and var-gene expression. After infection, waning immunity is also tracked and taken into account upon reinfection [44]. Individuals may remain asymptomatic, or develop a mild, clinical, or severe case of malaria. Due to the uncertain pathway from severe malaria to death, malaria-specific mortality was not enabled in this analysis.

Upon model initiation, all individuals are immunologically naive to malaria (though a few individuals are initialized as having an active infection). To generate a realistic immune profile, EMOD is typically run with no interventions for several decades, at which point the simulation state can be "frozen". From this starting point, a range of different intervention scenarios can be initialized.

Malaria Interventions EMOD offers a wide range of malaria interventions, in which the user can specify intervention timing, coverage, targeting, efficacy, and waning, among many other variables. Interventions function by altering default model probabilities– for example, giving an individual an insecticide-treated net would reduce their probability of being bitten by a mosquito indoors, while receiving antimalarial drugs after infection would alter intrahost immune parameters which, in turn, would lower the likelihood of a mosquito acquiring gametocytes and increase the rate of recovery from clinical disease.

Table 1.2 presents brief descriptions of the malaria interventions used in this analysis. See the EMOD documentation [43] for more.

1.3 Results

Presented here are the results of the example configuration.

1.3.1 Covariate Scaling, Dimensionality Reduction, and Clustering

Before rescaling, synoptic ERA5 air temperature values ranged from 2.2 to 39.0 degrees Celsius, with a median value of 24.9 degrees and an Interquartile Range (IQR)

Intervention	Details
Insecticide-Treated Nets (ITNs)	Distributed annually on January 1st; Initial blocking 90%, decaying exponentially with a 2-year half-life; Initial killing 60%, decaying exponentially with a 4-year half-life; Nets are discarded with exponential decay and a 6-month half-life; A randomly-selected 10% of the population never receives a net; Everyone who owns a net uses it every night.
Indoor Residual Spraying (IRS)	Distributed annually on January 1st; Initial killing 60%, decaying exponentially with a 6-month half-life; No correlation between IRS and ITN coverage.
Artemether-Lumefantrine Case Management (AL CM)	Constant throughout simulation; "Coverage" refers to the probability of seeking care given a clinical case; Upon deciding to seek care, per-day probability of care-seeking is 0.15; No age dependence; Drugs clear infection and provide about two weeks of additional protection.
Dihydro-artemisinin-Piperaquine Case Management (DP CM)	Constant throughout simulation; Same as AL case management, but with 1 month of protection after use.
Pre-Erythrocytic Vaccine (PEV)	Constant throughout simulation; 90% acquisition blocking, decaying exponentially; 6- and 12- month decay half-lives; Distributed to infants upon reaching 6 months of age.
Transmission-Blocking Vaccine (TBV)	Distributed annually on January 1st; 90% transmission blocking, decaying exponentially; 6- and 12- month decay half-lives; Distributed to adults age 15-49.
Attractive Targeted Sugar Baits (ATSB)	Distributed twice per year on January 1st and July 1st; Reported in terms of per-feed mortality rate, not coverage; 6-month box duration of efficacy.
Monoclonal Antibodies (mAbs)	Distributed annually on January 1st; Simulated via a PEV (see parameters above); Distributed to adults age 15-49; 3-month box duration of protection.

Table 1.2: EMOD intervention descriptions.

from 22.4 to 27.0. ERA5 precipitation ranged from 0 to 3,058 mm/month, with a median value of 40.6 and an IQR from 4.9 to 130.8. Precipitation values were capped at the 99th percentile cutoff of 366.9 mm/month. Relative mosquito abundance values ranged from 0 to 1, with continent-wide proportions of 43.0% *arabiensis*, 36.8% *gambiae*, and 20.2% *funestus* among those pixels with mosquitoes. Figure 1.1 shows covariate distributions before and after rescaling. Both precipitation and mosquito vector abundance show peaks near zero and one, while temperature peaks most strongly in the center of the distribution.

Figure 1.2 shows the variance explained by the first five singular vectors in the SVD procedure. The first three singular vectors accounted for 95.3% of all variance and were retained for the clustering analysis. Figure 1.3 shows the elbow plot for the k-means procedure. It shows that, while a k of ten captures over 85% of total data variance as between-cluster variance, a slightly higher cluster count would have covered over 90% of this ratio. Figure 1.4 shows a snapshot of the reprojected data points in 3D space, colored according to the k-means results. Black dots represent true cluster centroids, while gray dots indicate representative sites. The shape of this dataset is roughly circular or toroidal, with those sites that have appreciable numbers of mosquitoes and precipitation comprising the edges of the circle and only the low-precipitation, low-mosquito-count lime green vector clustering near the origin. Visual examination shows that the ten-cluster k-means differentiates several distinct groups, but may combine some groups that are visually fairly distinct, such as the two lobes of the turquoise cluster (Cluster 4).

1.3.2 Archetype Creation and Site Selection

Figure 1.5 shows archetype maps over a range of cluster counts from 3 to 14. Across all cluster counts, archetypes tend toward geographic contiguity. Given that the

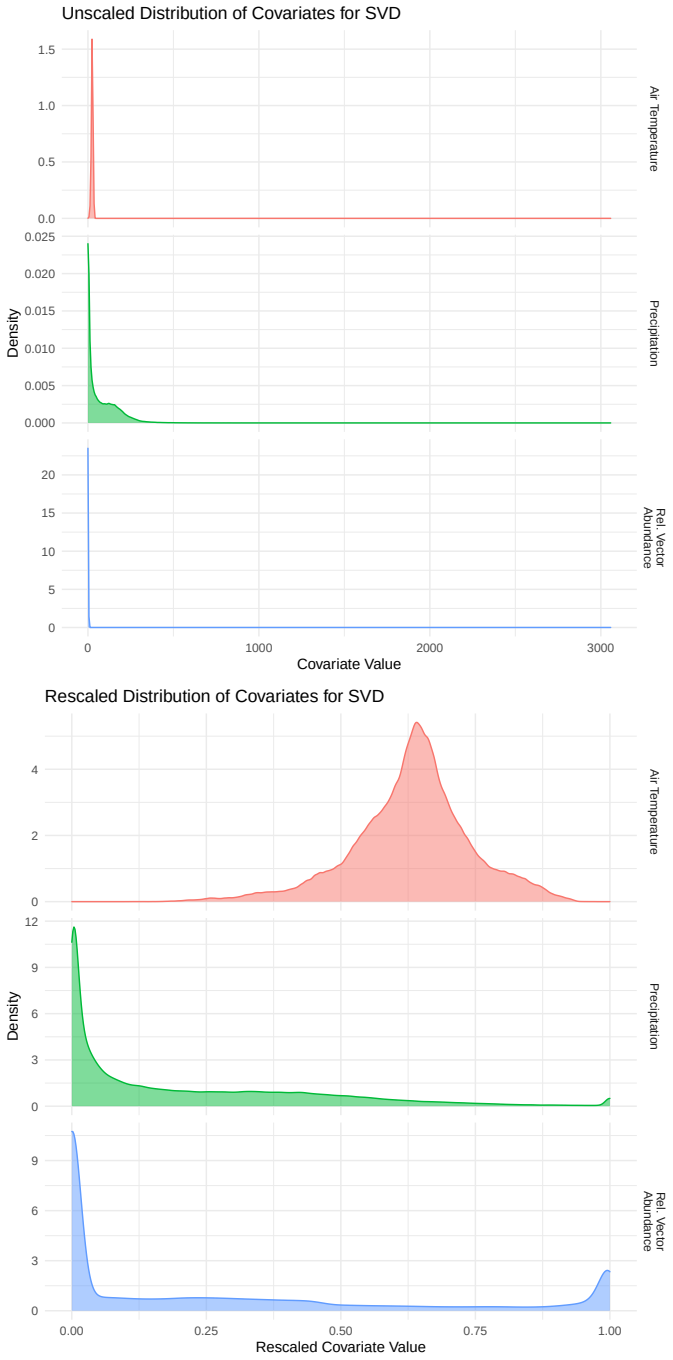


Figure 1.1: Distribution of covariates for Singular Value Decomposition (SVD) across all pixels before and after rescaling. Synoptic monthly mean temperature (degrees C) and precipitation (mm/month) are from the ERA5 project. Relative vector abundance, a static metric, tracks the proportion of *Anopheles arabiensis*, *funestus*, and *gambiae* in each pixel.

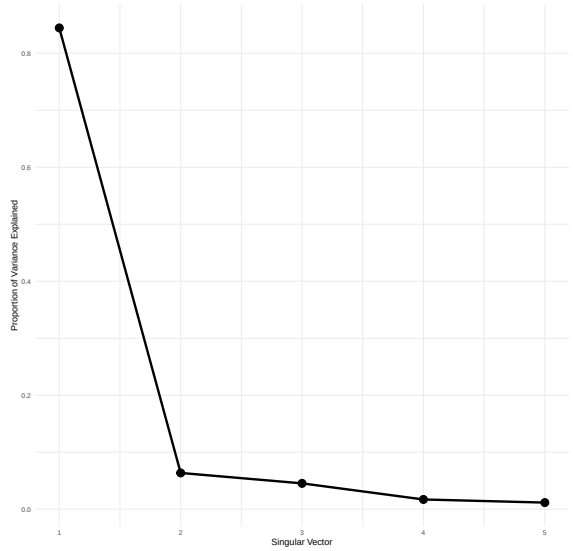


Figure 1.2: Proportion of variance explained by the first five singular vectors after Singular Value Decomposition (SVD). The first three vectors were retained for clustering analysis.

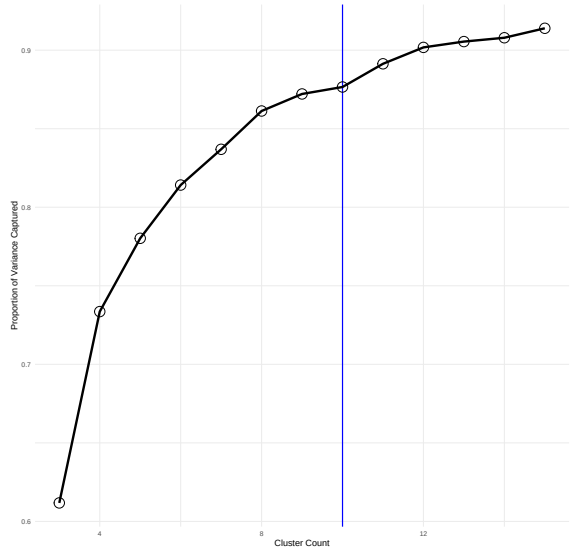


Figure 1.3: K-means elbow plot for k between 3 and 15. The x-axis shows cluster count, while the y-axis shows what proportion of total data variance is captured by between-cluster variance in each setting. A cluster count of ten (vertical blue line) was chosen for further analysis.



Figure 1.4: Top and side view of pixel values reprojected onto the first three singular vectors, after k-means clustering with a k of ten. Colors refer to different clusters and match subsequent ten-cluster plots. Black dots indicate true cluster centroids, while dark grey dots show procedurally-selected representative sites.

clustering covariates themselves demonstrate strong spatial autocorrelation, this result is perhaps unsurprising, but lends algorithmic credence to historical strategies of heuristically grouping regions by climate or seasonality. Also notable is the stability of archetype affiliation across different cluster counts. The three-cluster map identifies a strongly seasonal Sahelian/coastal band across west Africa, a band in central Africa with opposite seasonality to the first, and more arid regions to the north and south. Adding clusters allows separation of the northern and southern arid regions, followed by increasingly distinct latitudinal bands in west Africa and a stable boundary between central Democratic Republic of the Congo (DRC) and northern Zambia and Mozambique. Lake Victoria, Rwanda, and Burundi are also reliably grouped together and distinct from their immediate surroundings, usually matched to an archetype further south. The southern tip of Madagascar is reliably grouped with southern Mozambique, while the rest of the island joins the same archetype as northern Mozambique and Zambia. The eastern coastline of Madagascar sometimes joins the archetype dominated by southern DRC. The ten-cluster map was selected for the example configuration to strike a heuristic balance between variance explained in the elbow plot (Figure 1.3) and the communicative value of showing archetypes with visually distinct seasonality and mosquito mixes.

Figure 1.6 shows a map of the ten archetypes used in the example configuration, along with the associated temperature and precipitation time series and relative abundance of different mosquito species. In the time series, solid colored lines represent the median across the archetype, shaded areas indicate the interquartile range, and dotted lines indicate the 95% variance interval. Solid black lines represent the climate values of the representative site for each archetype, also indicated as black crosses on the map. Section 1.5 shows similar maps for all cluster counts tested. The ten-archetype setting captures a detailed range of different climate modalities on the continent. Three latitudinal bands across the Sahara to the coast show seasonal precipitation of increasing magnitude and duration (magenta, turquoise, and

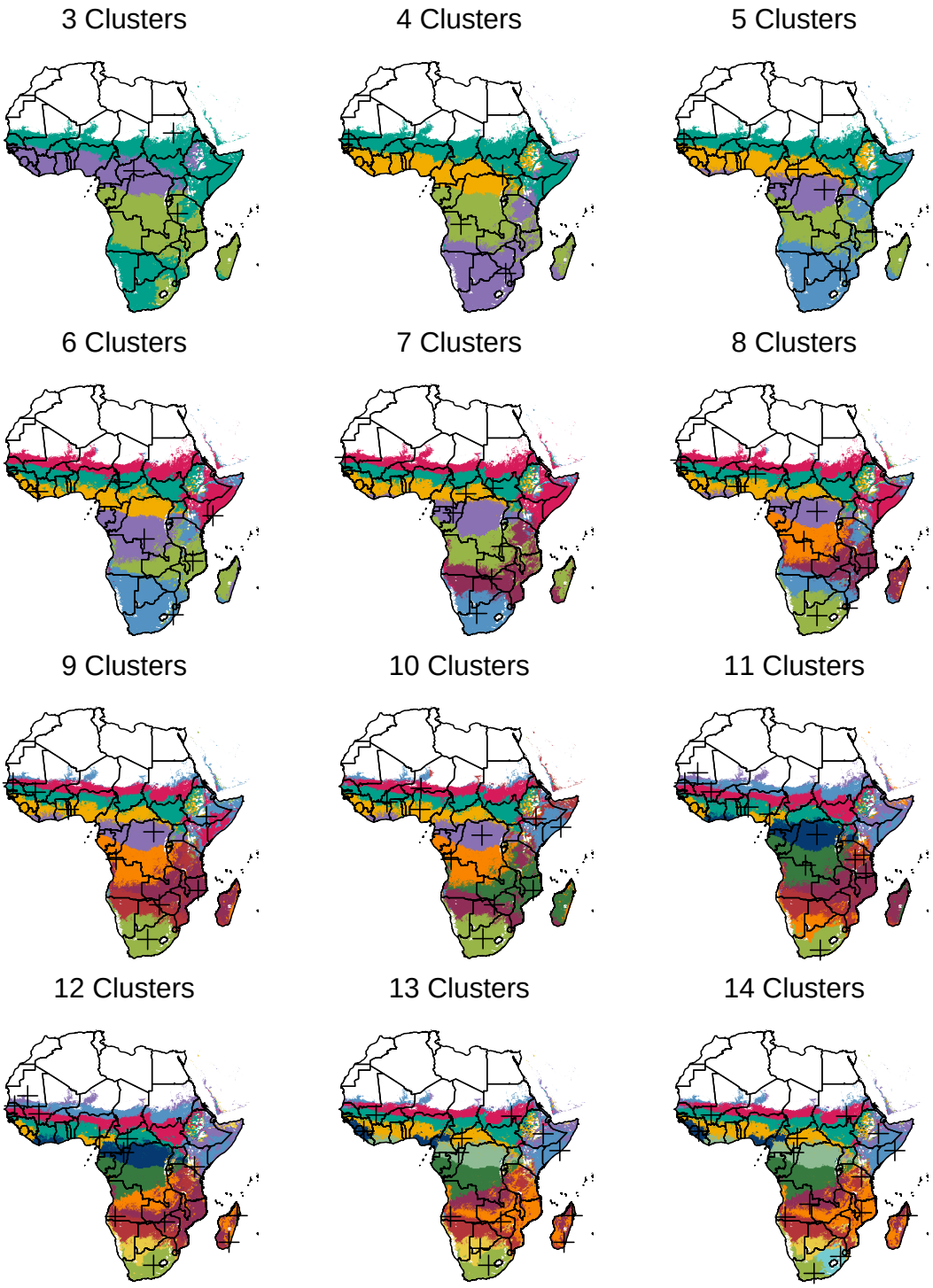


Figure 1.5: Cluster maps for k of 3 to 14. Black crosses indicate procedurally-generated representative sites (the pixels closest to cluster centroids).

gold). Relative *arabiensis* proportions decline, and *gambiae* proportions increase, along the same gradient. These latitudinal bands continue down the western side of the continent, with a bimodal *gambiae*-dominated archetype in northern DRC (purple), a strongly seasonal *gambiae*-dominated archetype in southern DRC (orange), and a more gently seasonal *arabiensis* and *funestus*-dominated archetype across northern Namibia and Botswana into southern Zimbabwe and Mozambique (plum). Archetypes are least contiguous along the continental divide, with Ethiopia, Uganda, and much of western Kenya showing a mix of archetypes. This effect is likely driven by the complex and highly varied topography of the Great Rift Valley, which introduces sharper gradients of temperature, precipitation, and mosquito species than more smoothly-varying landscapes elsewhere on the continent. The horn of Africa is classified into two bimodal archetypes dominated by *arabiensis*, one with a substantial fraction of *funestus* (maroon) and one without (blue). Zambia, northern Mozambique, southern Tanzania, and most of Madagascar comprise a strongly seasonal archetype with a near-even mix of all three vectors (dark green). A broad area of central Tanzania is associated with the lower-precipitation, *arabiensis*-dominated southern band rather than any of the archetypes it borders (plum). The lime green archetype in southern Africa is distinguished by having few to no mosquitoes, and is excluded from simulation analysis. With the exception of the two representative sites in the horn of Africa, which display considerably higher precipitation than their archetypes' median values, representative site time series line up well with the time series of the archetype medians. These two clusters have fewer members and higher variability than many of the others, which may explain this effect. Table 1.3 describes the geography and mosquito characteristics of each representative site.

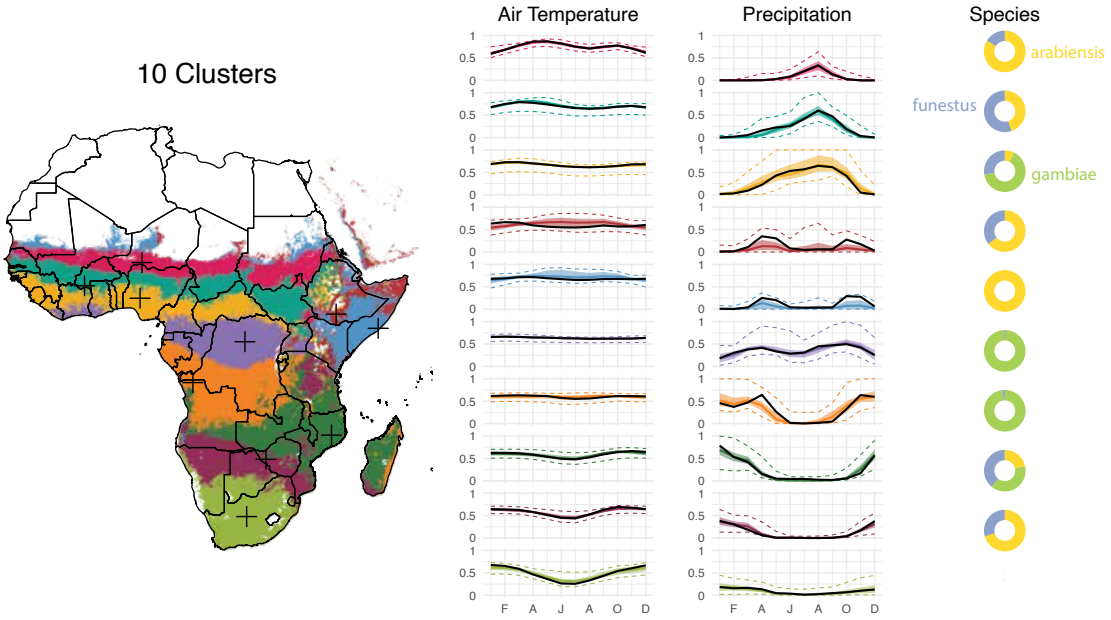


Figure 1.6: Maps and time series for the ten-site setting. In the time series, solid colored lines represent the median across the archetype, shaded areas indicate the interquartile range, and colored dotted lines indicate the 95% variance interval. Solid black lines represent the procedurally-selected climate values of the representative site for each archetype, also indicated as black crosses on the map. Doughnut plots show the relative vector abundance of the representative sites.

1.3.3 Simulation Outputs and Intervention Maps

Figure 1.7 shows example impact curves for each archetype under different intervention scenarios, as well as their translation into maps of predicted impact. The impact curves show pre-intervention malaria prevalence on the x-axis and malaria prevalence after three years of interventions on the y-axis. The line of equivalence, shown in grey, indicates no change between initial and final prevalence, while the colored curves show simulation results across the range of transmission intensities. More effective intervention packages have curves that swing farther toward the lower right corner of the plots.

The first scenario, upper left map and dot-dashed lines, does not have interventions and therefore recreates the baseline 2000 prevalence map. The second, upper right map and dotted lines, shows a complex intervention mix of annual Insecticide-Treated Net (ITN) distributions with moderate coverage (40%), annual Indoor Residual Spraying (IRS) campaigns with 20% coverage, and low access and use of Artemether-Lumefantrine Case Management (AL CM), with 20% of clinical cases receiving treatment. The third, lower left map and solid lines, shows three years of high antimalarial access and use (80% of clinical cases receiving treatment), but no other interventions. The fourth, lower right map and dashed lines, shows a scenario in which the only intervention is a hypothetical Attractive Targeted Sugar Bait (ATSB) with a 3% mortality rate. For more intervention details see Table 1.2.

In all cases, initial prevalence is the primary driver of intervention impact, but the importance of accounting for archetype is highlighted in comparing scenarios. For example, the upper and lower right maps generally produce similar results despite consisting of non-overlapping interventions. However, difference between the two are clear especially in the plum-colored archetype, in which Angola and southern Zambia have different elimination outcomes between the two exercises and southern

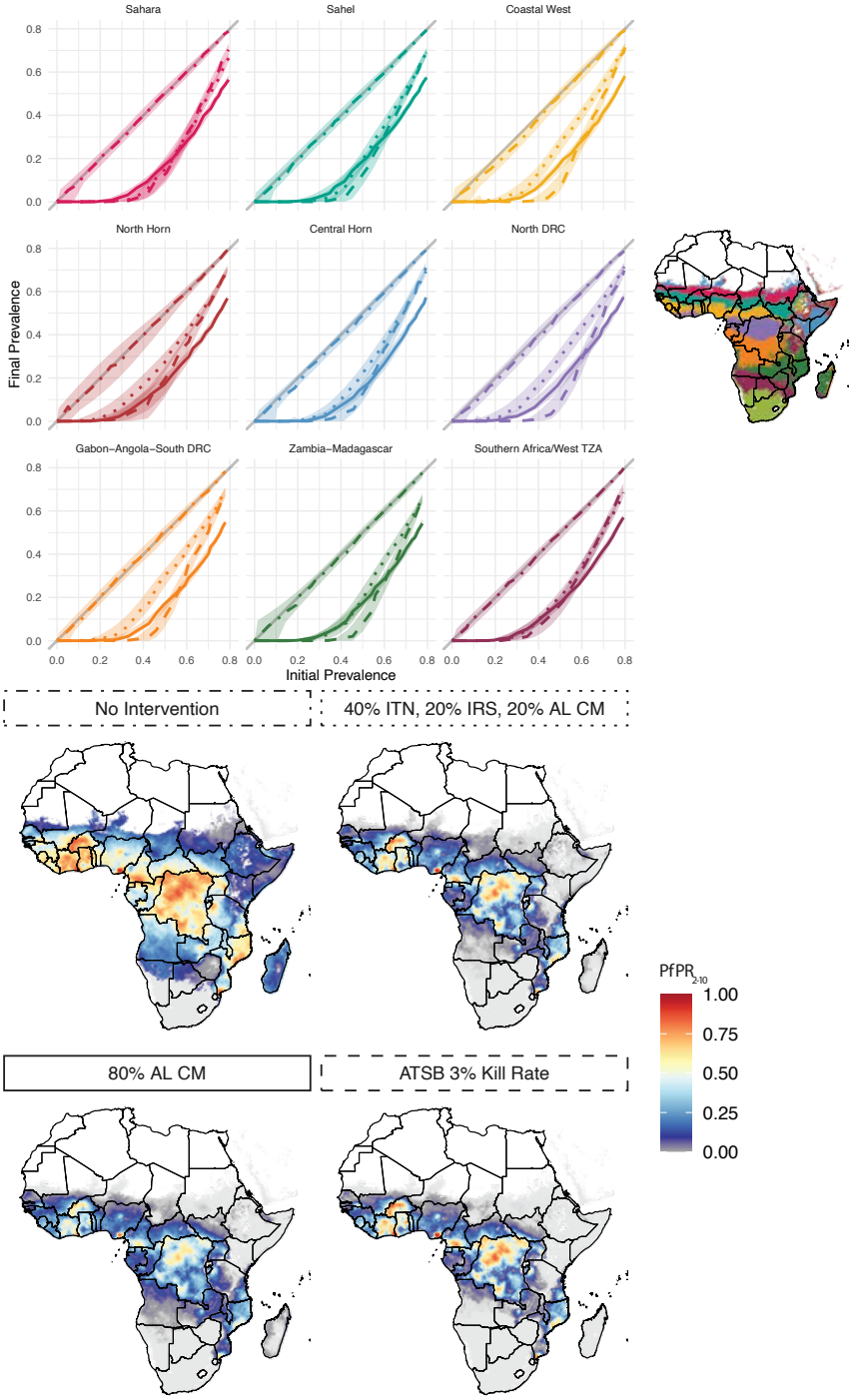


Figure 1.7: Example intervention impact results for four of the 152 interventions tested. The line plots (top) show intervention impact curves disaggregated by archetype, with archetypes colored according to the map (top right). The four prevalence maps (bottom) show hypothetical *Plasmodium falciparum* parasite rate among 2-10 year olds ($PfPR_{2\text{-}10}$) for each scenario. These maps were created by starting from a baseline map of 2000 as a proxy for a landscape without malaria interventions. For each initial prevalence value in the baseline map, a new prevalence value was found using the lines in the top set of plots. The linetype of subplot borders in the lower plots matches the linetype of interventions in the upper plots. For more detail about interventions see Table 1.2.

Mozambique performs better under an ATSB approach than a complex intervention mix. The antimalarials-only map differs considerably from both of the other intervention packages. In particular, it is more effective than either of the others at high baseline prevalence levels, but often comparable or less effective at low baseline prevalence. This generates a map with a similar elimination landscape, but many fewer hotspots, than the other two. Plots and maps of all 152 interventions are available under the "10-Site Setting" label at https://institutefordiseasemodeling.github.io/archetypes-intervention-impact/.

Each archetype, because of its unique mix of vectors, possesses a different percentage of mosquito bites that occur indoors as opposed to outdoors. It is important for this model framework to capture the differential impact of indoor-biting-targeted interventions in settings with different indoor biting rates. Figure 1.8 demonstrates this capability. Here, each line represents an archetype, colored according to its indoor biting percentage as determined by its mosquito species mix. The left panel shows an intervention setting with high coverage of insecticide-treated nets and indoor residual spraying, both of which target mosquitoes indoors. This intervention package, as expected, is much more effective in areas with a higher percentage of bites occurring indoors. The right panel demonstrates the inverse property: under an intervention package that includes only anti-malarial drugs and therefore does not directly target mosquitoes at all, there is no differential impact across indoor biting intensities. Table 1.3 describes the mosquito characteristics of each representative site.

1.3.4 Site Selection Sensitivity Analysis

Figure 1.9 shows the location of all 100 sensitivity sites. With one exception, the sensitivity analysis showed that the representative sites appropriately capture varia-

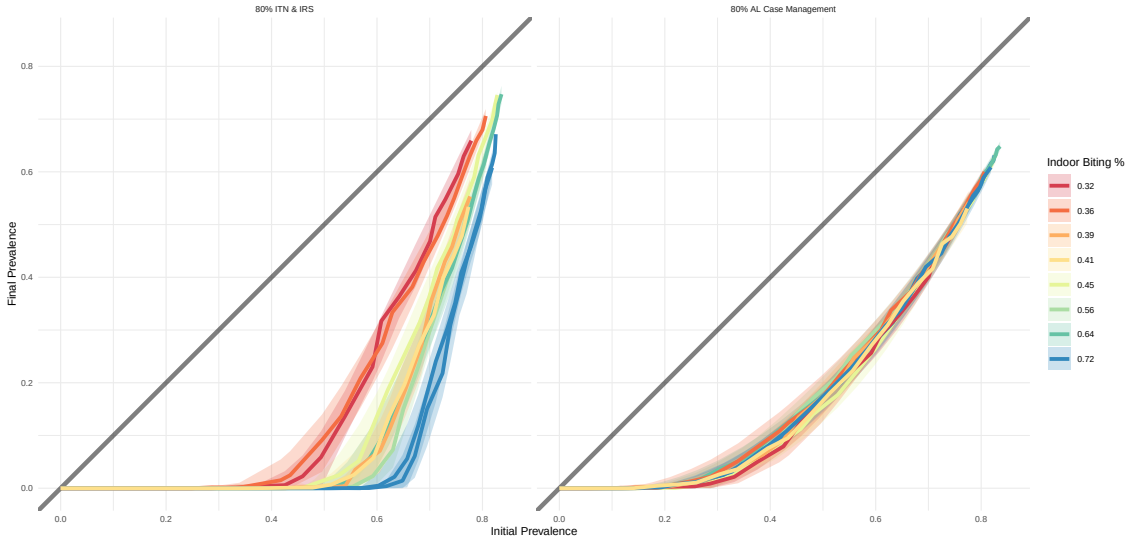


Figure 1.8: Sensitivity of sites to biting intensity for vector control vs non-vector-control interventions. In both plots, each line represents an archetype's representative site, colored according to its indoor biting percentage as determined by its mosquito species mix. The left panel shows an intervention setting with 80% coverage of Insecticide-Treated Net (ITN) and Indoor Residual Spraying (IRS), both of which target mosquitoes indoors. This intervention package, as expected, is much more effective in areas with a higher percentage of bites occurring indoors. The right panel demonstrates the inverse property: under an intervention package that includes only anti-malarial drugs and therefore does not directly target mosquitoes at all, there is no differential impact across indoor biting intensities.

tion within archetypes. Figure 1.10 shows an example intervention package, with the impact curve of each archetype's representative site plotted in color and the curves of the ten randomly-selected sites shown in black. Shaded areas represent variation across the ten random seeds run for each site. If the representative sites were effective proxies for their archetypes, the colored lines and shaded areas would cover all or most of the black lines and shaded areas. While there are a few archetypes in which the representative sites' variation is slightly too narrow, especially the yellow and blue curves, in general the representative sites capture both the shape and the variation across the sensitivity sites. The one exception is the plum-colored archetype, in which the representative site curve hews closely to seven of the ten sensitivity sites, but completely misses the curves of the three others. This is perhaps unsurprising since this archetype is the least spatially contiguous, consisting of one band across Namibia, Botswana, Zimbabwe, and Mozambique, and a second vertical swath of southern Kenya and much of Tanzania. The representative site for this archetype is in Zimbabwe, while the sensitivity sites range in location from central Tanzania to deep in the Namibian desert. That all of these sites landed in the same archetype suggest a benefit to including more archetypes in future analyses, or setting constraints on spatial contiguity of archetypes. Similar plots for all 152 intervention packages are available under the "Sensitivity Analysis" label at https://institutefordiseasemodeling.github.io/archetypes-intervention-impact/.

1.4 Discussion

Mechanistic models are versatile tools for malaria intervention planning, but deploying them over large geographic areas at a useful spatial resolution poses both theoretical and computational challenges. This chapter introduces a flexible and

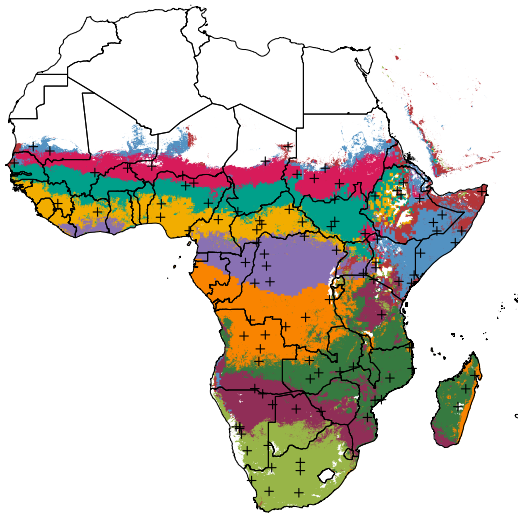


Figure 1.9: Sensitivity analysis sites. Black crosses indicate the location of all 100 representative sites used for sensitivity analysis.

Site #	**Description**	***arabiensis* Proportion**	***funestus* Proportion**	***gambiae* Proportion**	**Anthropophily (%)**	**Endophily (%)**	**Total Indoor Biting (%)**
1	Lime, Southern Africa	0	0	0	0	0	0
2	Teal, Sahel	0.45	0.55	0	65	70	45
3	Purple, northern DRC and coastal west Africa	0	0.01	0.99	85	85	72
4	Gold, west Africa	0.07	0.28	0.65	78	82	64
5	Blue, horn of Africa	1	0	0	65	50	32
6	Magenta, Saharan border	0.84	0.16	0	65	56	36
7	Plum, southern Africa and central Tanzania	0.71	0.29	0	65	60	39
8	Orange, southern DRC to Gabon	0	0.02	0.98	85	85	72
9	Maroon, northern Somalia	0.64	0.36	0	65	62	41
10	Green, Zambia, Mozambique, Malawi, and Madagascar	0.22	0.39	0.39	73	78	56

Table 1.3: EMOD Site Descriptions. DRC: Democratic Republic of the Congo.

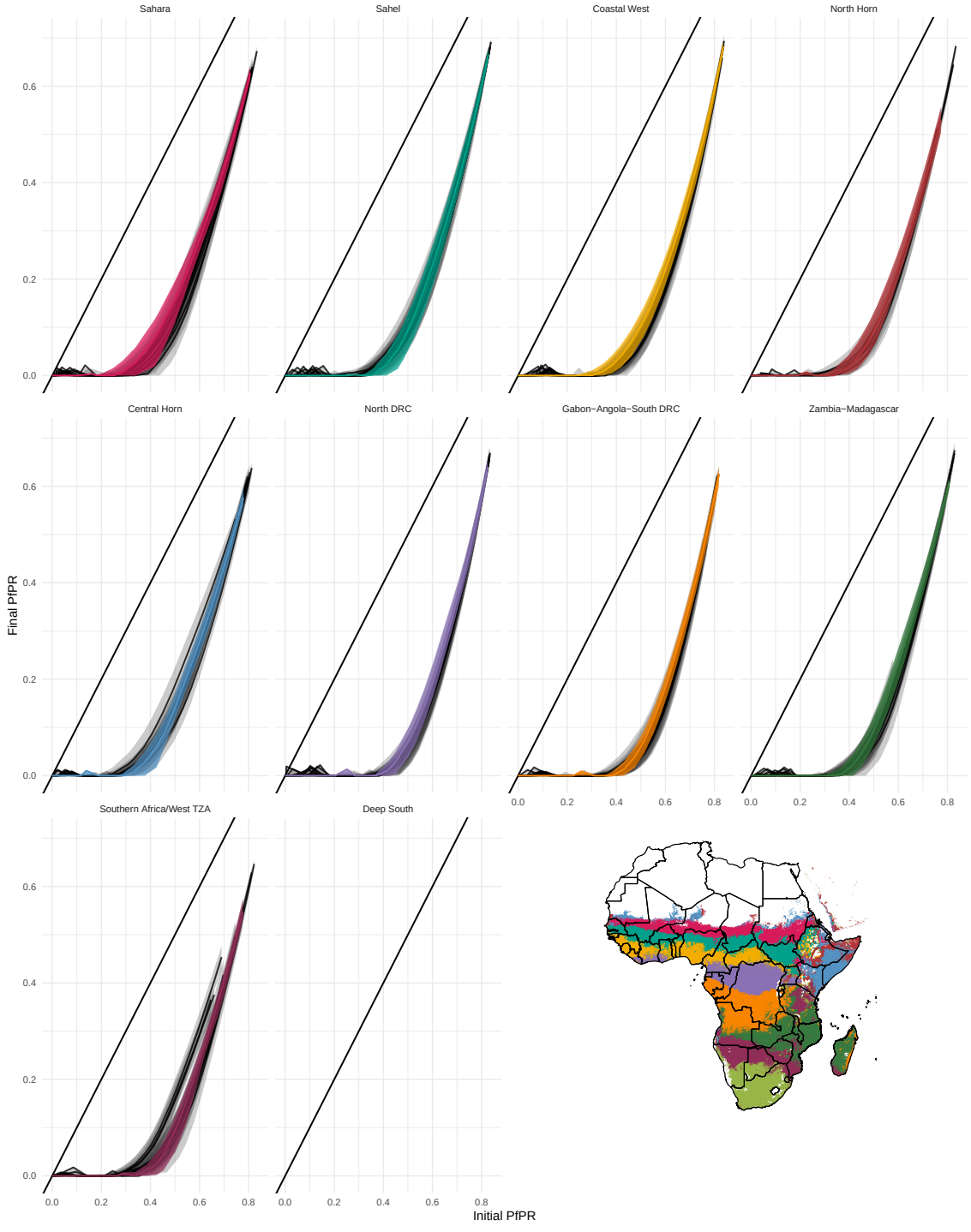


Figure 1.10: Sensitivity analysis example intervention curves for an intervention package of 40% Insecticide-Treated Net (ITN) coverage, 20% Indoor Residual Spraying (IRS) coverage, and 20% coverage with Artemether-Lumefantrine Case Management (AL CM). The impact curve of each archetype's representative site is plotted in color and the curves of the ten randomly-selected sites are shown in black. Shaded areas represent variation across the ten random seeds run for each site. If the representative sites were effective proxies for their archetypes, the colored lines and shaded areas would cover all or most of the black lines and shaded areas.

customizable archetypes-based framework that harnesses the richness of available spatial data and spatiotemporal modeling products to assess potential intervention impact across a range of settings, and discusses key decision points to consider when constructing such a framework. It includes a detailed example of how the framework might be configured. This example configuration demonstrates that dimensionality reduction and clustering can identify meaningfully different environmental archetypes, that the mechanistic model behaves as expected in relation to intervention effect and mosquito bionomics, that the creation of new maps showing intervention impact can highlight interventions and areas of interest, and that the representative sites selected can appropriately represent each archetype as a whole. The demonstrative analysis used the EMOD mechanistic model and malaria risk surfaces from MAP, but this methodology could just as easily be applied with different modeling software or malaria risk estimates. The choice of covariates upon which to cluster is likewise arbitrarily modifiable, noting that the mechanistic model used should have the capacity to accurately reflect any covariate variation that distinguishes archetypes from one another.

An archetypes-based approach can be useful for any project in which there is interest in the differential effects of a disease model under a range of initial conditions. Historically in malaria, such analyses either used idealized "characteristic" settings that did not directly represent real-world locations [28], or tested an extremely wide combinatoric collection of possible initial conditions [4]. The first approach is simple and straightforward, but suffers from arbitrary selection and the inability to claim that the sites shown are actually representative of other areas. The second approach, while thorough, requires an unnecessary computational burden. An archetypes-based strategy provides a useful middle ground, providing a relatively streamlined computational infrastructure for projects that require national or continental coverage, while also offering a useful suite of example settings that can easily be used for any type of exploratory simulation analysis.

The example configuration shows the utility of an archetypes framework for scenario-based analyses that are not calibrated to field data, but the framework has been extended to more data-rich use cases. As part of the recent High Burden High Impact (HBHI) collaborations between modeling groups, the World Health Organization (WHO), and malaria control programs, researchers from both IDM and Northwestern University utilized a covariate-based dimensionality-reduction and clustering strategy to generate subnational archetypes in Burkina Faso and Nigeria. These archetypes then helped inform a rigorous admin-level calibration and stratification process to recommend targeted intervention strategies in upcoming years in these countries [45, 46]. This general framework is now being extended to Burundi, Sierra Leone, and others. In addition, it has repeatedly proven useful to have a diverse range of archetypal models pre-configured when stakeholders ask policy questions that are not geographically specific but may have different answers in different settings. For example, the IDM team was recently asked by stakeholders at the Bill and Melinda Gates Foundation (BMGF) to run simulations assessing the potential epidemiological impact of deploying more durable insecticide-treated nets in Africa. For this analysis, a full recreation of continent-level maps was not required, but modelers were able to quickly and easily use the archetypes framework to run simulations over a range of meaningful seasonal and entomological settings to assess how the answer to this question might differ across the continent.

While identifying different malaria seasonality modalities was not the primary goal of this work, the archetype selection process consistently identified spatial groupings with similar seasonality to heuristically identified seasonality profiles [4, 47]. This observation lends confidence to the effectiveness of both the archetypes framework and other seasonality detection methods.

An archetypes-based approach to intervention impact planning has several limitations. When utilized without fitting to field data, as shown in this document, this

method is useful as an intuition-building tool, but should not be used to inform specific decisions in particular geographic locations. While the ability to project results from the analysis to an arbitrarily fine spatial scale may be useful for highlighting heterogeneities, such results might convey an unintended sense of confidence in the sensitivity and specificity of results. It is best to consider this framework as a somewhat principled method for obtaining informative prior distributions, rather than as a way of generating quantitatively rigorous model results. However, as described above in the HBHI framework, these concerns fade when the strategy is utilized as a precursor to more formal model fitting.

The example configuration specifically has additional limitations. While the selected covariates cover many environmental model sensitivities, no human-centered or malaria intervention history covariates were included. While some covariate standardization was performed, a limited number of rescaling or standardizing approaches were conducted. Similarly, while common heuristics were utilized to select singular vector and cluster counts in SVD and k-means, a full and formal sensitivity analysis of these heuristics has yet to be conducted. These would include running k-means with more or fewer singular vectors, and repeating the clustering analysis with k-fold data holdouts. Both of these tests would check how cluster membership changes in response to these varied initial conditions, especially for pixels on the borders between archetypes. However, the robustness of cluster assignment between different k-values provides some confidence that such sensitivity checks would support the choices made in this analysis. While the EMOD modeling software has been extensively vetted and tested, it has not undergone a parameter-by-parameter sensitivity analysis to conclusively determine which variables are most robust to model changes. Because the example configuration focused on hypothetical scenarios rather than prediction or projection, uncertainty was not considered. A full uncertainty propagation strategy would begin with covariate uncertainty and reflect both mechanistic model parameter uncertainty and uncertainty in the prevalence

maps used as baseline metrics.

The example configuration additionally does not consider a number of important malaria-related factors, but could easily be extended to take them into account. These include the absence of insecticide resistance, human migration, historical interventions, and site-specific population demographics. Future work will also include model comparison exercises between the archetypes framework and models that have been finely calibrated to specific locations to assess what improvements can be made in archetype-level predictions (while continuing to acknowledge that superseding site-specific analysis is not and will not be the goal). The examples shown here are in Africa, but this framework can also be extended to malaria-endemic regions of Asia and the Americas.

This chapter presents a novel archetypes-based strategy for high-resolution, large-area malaria intervention impact assessment. Compared to similar approaches, it adds speed, computational efficiency, and interpretability to results, lends itself well to more detailed calibration-based approaches, and guides intuition in data-sparse settings. This chapter focused on the methodological theory and mechanics of the archetypes framework, utilizing an example configuration as an example. Chapter 2 fleshes out a case study of this framework in more detail, providing context and comparing the results presented here to an earlier and more heuristics-based version of the framework.

1.5 Additional Plots

The plots below show maps and covariate summaries for cluster counts of three to 14. In the time series, solid colored lines represent the median across the archetype, shaded areas indicate the interquartile range, and dotted lines indicate the 95%

variance interval. Solid black lines represent the climate values of the representative site for each archetype, also indicated as black crosses on the map. Doughnut plots show the relative vector abundance of the representative sites. Absent doughnuts indicate no mosquitoes in that site.

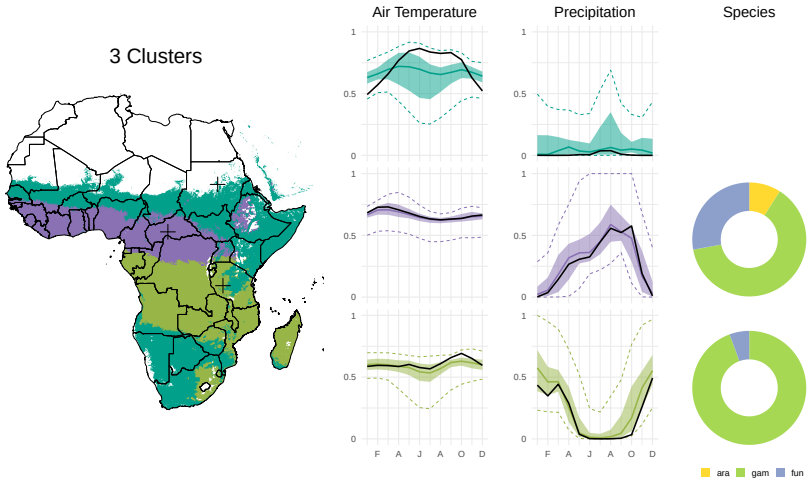


Figure 1.11: Maps and variable summaries for a cluster count of 3.

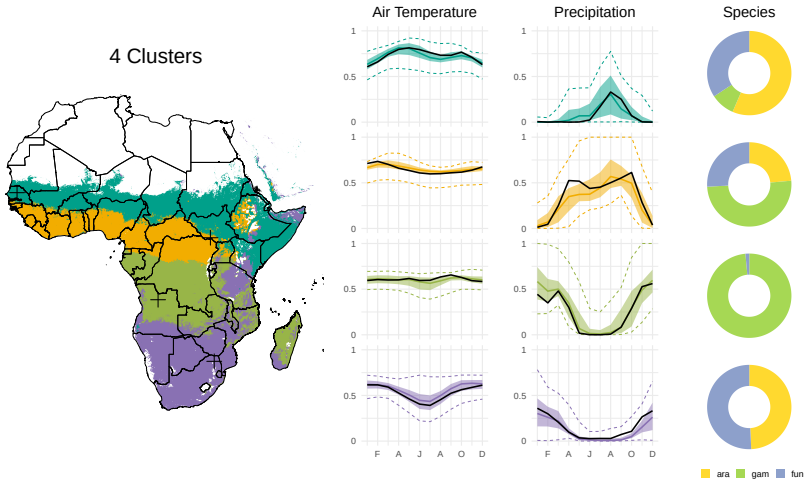


Figure 1.12: Maps and variable summaries for a cluster count of 4.

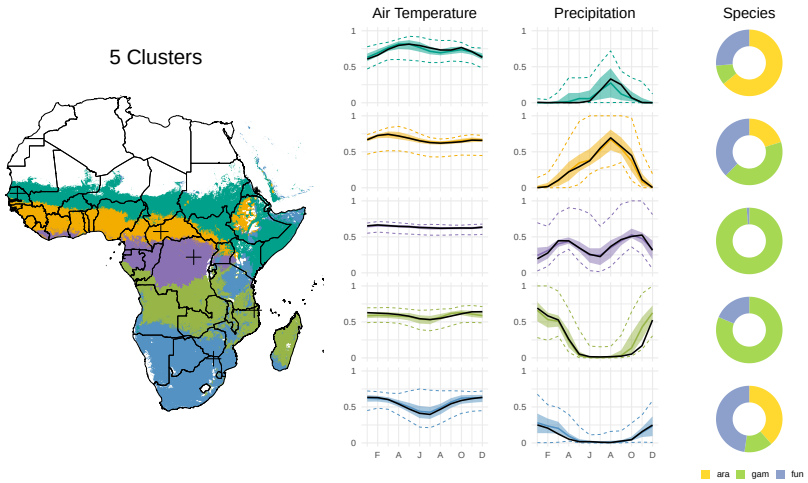


Figure 1.13: Maps and variable summaries for a cluster count of 5.

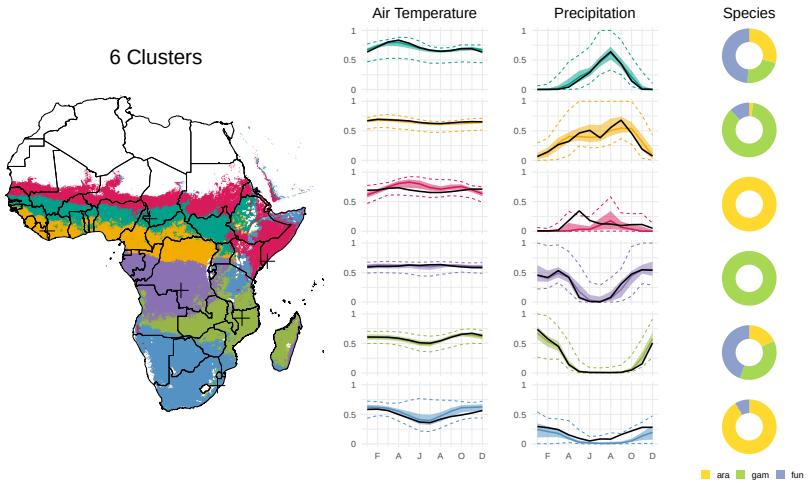


Figure 1.14: Maps and variable summaries for a cluster count of 6.

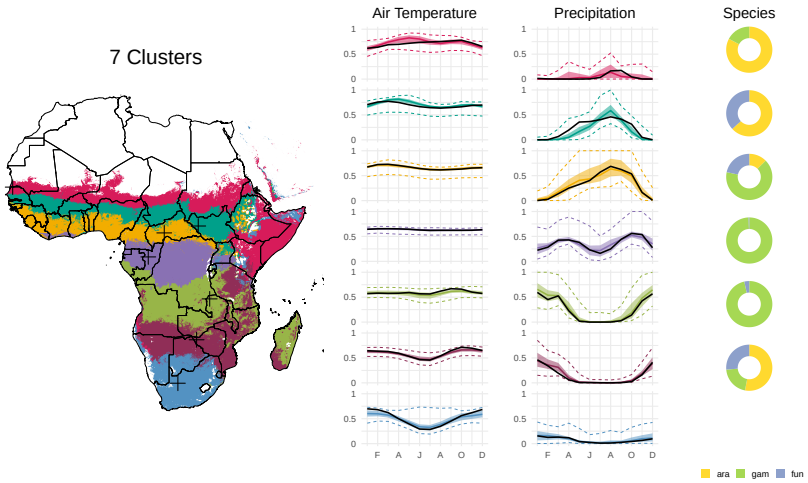


Figure 1.15: Maps and variable summaries for a cluster count of 7.

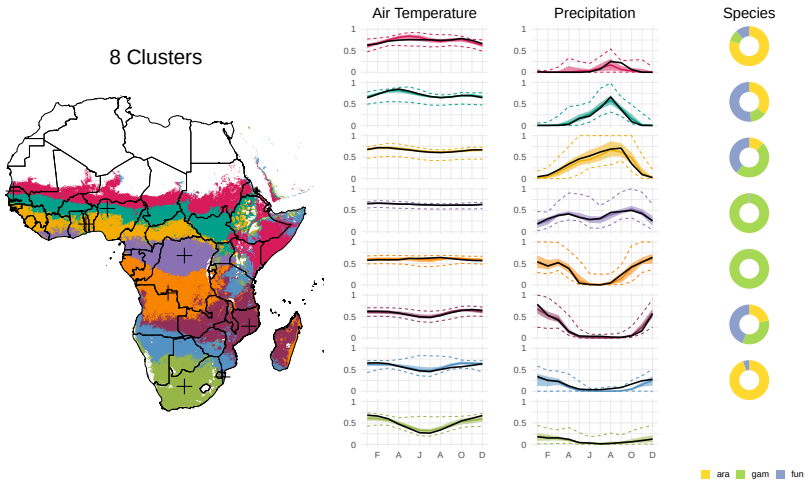


Figure 1.16: Maps and variable summaries for a cluster count of 8.

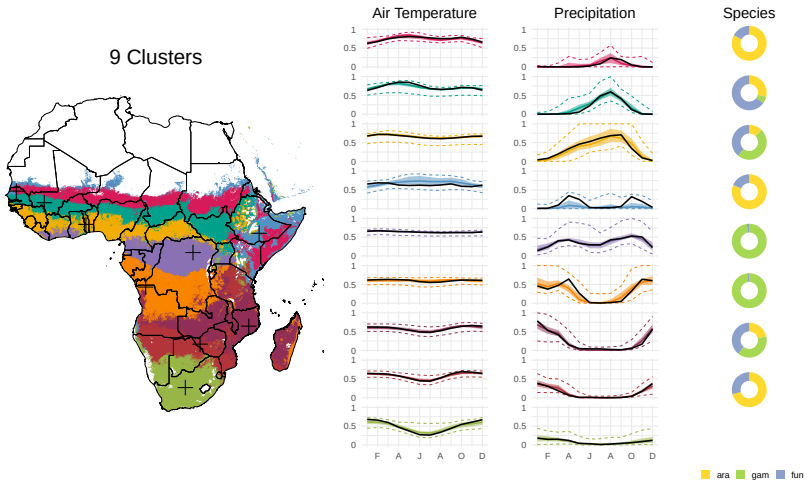


Figure 1.17: Maps and variable summaries for a cluster count of 9.

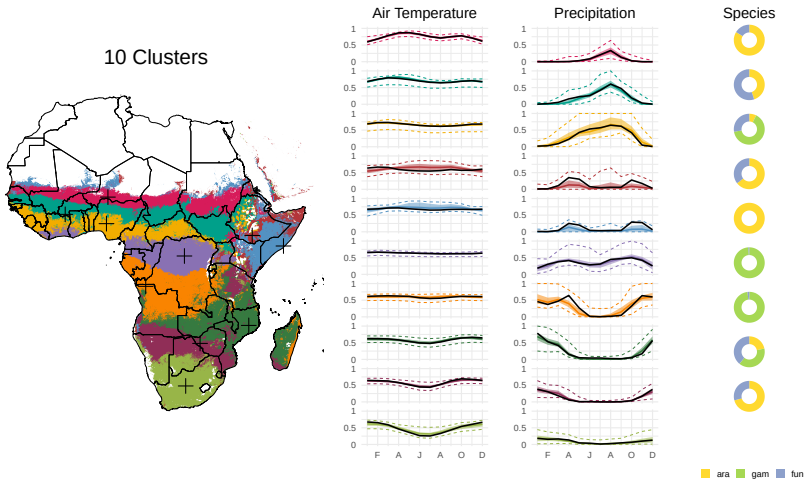


Figure 1.18: Maps and variable summaries for a cluster count of 10.

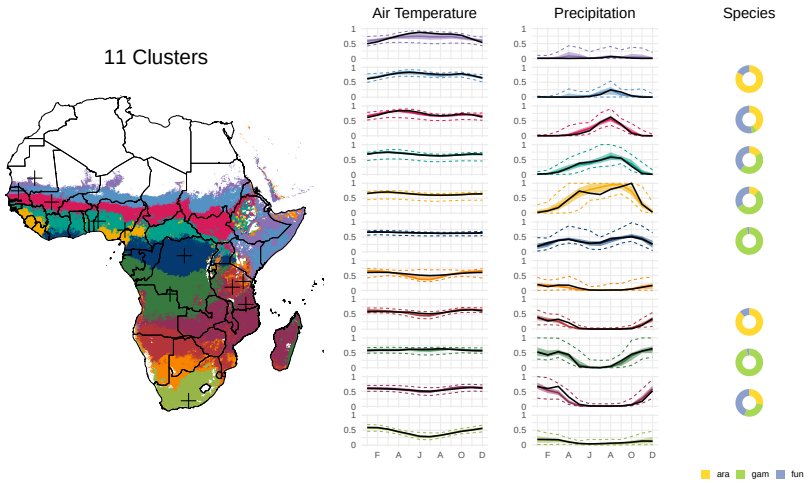


Figure 1.19: Maps and variable summaries for a cluster count of 11.

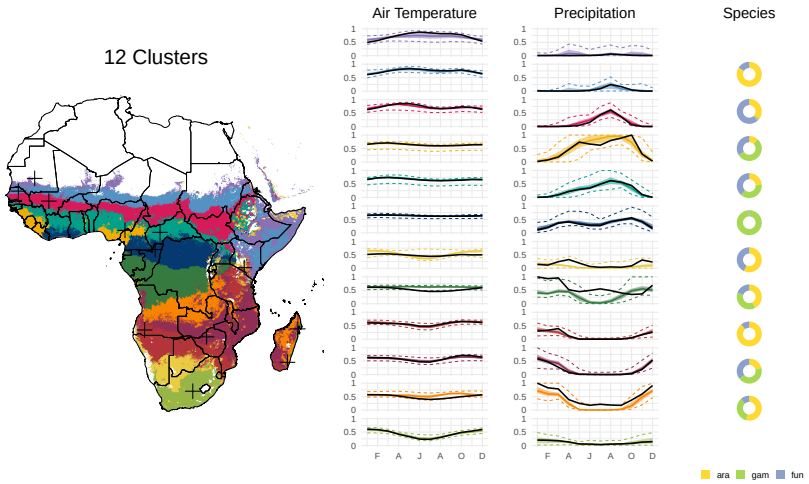


Figure 1.20: Maps and variable summaries for a cluster count of 12.

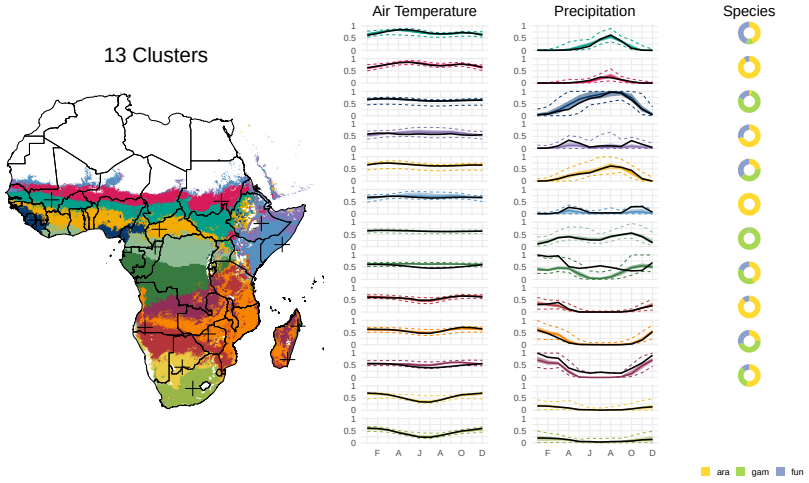


Figure 1.21: Maps and variable summaries for a cluster count of 13.

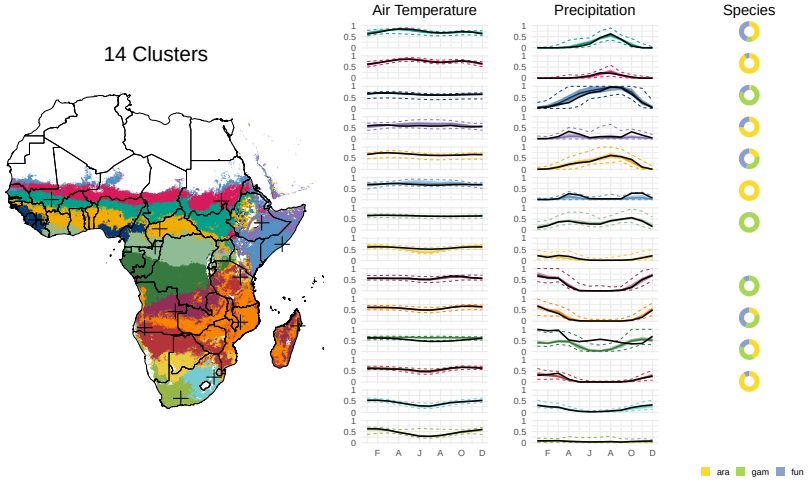


Figure 1.22: Maps and variable summaries for a cluster count of 14.

Bibliography

1. Packard, R. M. *The Making of a Tropical Disease: A Short History of Malaria* ISBN: 9781421403960 (Johns Hopkins University Press, Baltimore, 2011).
2. Nájera, J. A., González-Silva, M. & Alonso, P. L. Some Lessons for the Future from the Global Malaria Eradication Programme (1955–1969). *PLOS Medicine* **8,** e1000412. ISSN: 1549-1676. https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000412 (2011).
3. Smith, D. L. *et al.* Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. *PLoS Pathogens* **8.** ISSN: 15537366 (2012).
4. Walker, P. G. T., Griffin, J. T., Ferguson, N. M. & Ghani, A. C. Estimating the most efficient allocation of interventions to achieve reductions in P.falciparum malaria burden and transmission in Africa. *The Lancet Global Health,* 1–11. ISSN: 2214109X. http://dx.doi.org/10.1016/S2214-109X(16)30073-0 (2015).
5. Crowell, V. *et al.* Modelling the cost-effectiveness of mass screening and treatment for reducing Plasmodium falciparum malaria burden. *Malaria journal* **12,** 4. ISSN: 1475-2875. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3544609%7B%5C&%7Dtool=pmcentrez%7B%5C&%7Drendertype=abstract (2013).
6. Stuckey, E. M., Smith, T. a. & Chitnis, N. Estimating malaria transmission through mathematical models. *Trends in Parasitology* **29,** 477–482. ISSN: 14714922. http://dx.doi.org/10.1016/j.pt.2013.08.001 (2013).
7. Eckhoff, P. a. A malaria transmission-directed model of mosquito life cycle and ecology. *Malaria Journal* **10,** 303. ISSN: 1475-2875. http://www.malariajournal.com/content/10/1/303 (2011).
8. Smith, N. R. *et al.* Agent-based models of malaria transmission: A systematic review. *Malaria Journal.* ISSN: 14752875 (2018).
9. Wakefield, J. C. & Morris, S. E. The Bayesian Modeling of Disease Risk in Relation to a Point Source. *https://doi.org/10.1198/016214501750332992* **96,** 77–91. https://www.tandfonline.com/doi/abs/10.1198/016214501750332992 (2011).
10. Wilson, K. & Wakefield, J. Pointless spatial modeling. *Biostatistics* **21,** e17–e32. ISSN: 1465-4644. https://academic.oup.com/biostatistics/article/21/2/e17/5092061 (2020).

11. Craig, M., Snow, R. & le Sueur, D. A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitology today (Personal ed.)* **15,** 105–111. ISSN: 0169-4758. https://pubmed.ncbi.nlm.nih.gov/10322323/ (1999).

12. Snow, R., Craig, M., Deichmann, U. & le Sueur, D. A preliminary continental risk map for malaria mortality among African children. *Parasitology today (Personal ed.)* **15,** 99–104. ISSN: 0169-4758. https://pubmed.ncbi.nlm.nih.gov/10322322/ (1999).

13. Kleinschmidt, I., Bagayoko, M., Clarke, G., Craig, M. & Sueur, D. L. A spatial statistical approach to malaria mapping. *International journal of epidemiology* **29,** 355–361. ISSN: 0300-5771. https://pubmed.ncbi.nlm.nih.gov/10817136/ (2000).

14. Tanser, F. C. & Le Sueur, D. The application of geographical information systems to important public health problems in Africa. *International Journal of Health Geographics*. http://www.ij-healthgeographics.com/content/1/1/4 (2002).

15. Hay, S. I. *et al.* Estimating the Global Clinical Burden of Plasmodium falciparum Malaria in 2007. *PLoS Medicine* **7.** /pmc/articles/PMC2885984/%20/pmc/articles/PMC2885984/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2885984/ (2010).

16. Gething, P. *et al.* A long neglected world malaria map: Plasmodium vivax endemicity in 2010. *PLoS neglected tropical diseases* **6.** ISSN: 1935-2735. https://pubmed.ncbi.nlm.nih.gov/22970336/ (2012).

17. Battle, K. E. *et al.* Mapping the global endemicity and clinical burden of Plasmodium vivax, 2000–17: a spatial and temporal modelling study. *The Lancet* **394,** 332–343. ISSN: 1474547X. http://dx.doi.org/10.1016/ (2019).

18. Weiss, D. J. *et al.* Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum, 2000–17: a spatial and temporal modelling study. *The Lancet* **394,** 322–331. ISSN: 1474547X (2019).

19. Hay, S. I. *et al.* Developing Global Maps of the Dominant Anopheles Vectors of Human Malaria. *PLoS Medicine* **7.** /pmc/articles/PMC2817710/%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2817710/ (2010).

20. Sinka, M. E. *et al.* Modelling the relative abundance of the primary African vectors of malaria before and after the implementation of indoor, insecticide-based vector control. *Malaria Journal* **15,** 142. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/s12936-016-1187-8 (2016).

21. Bhatt, S. *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* **526,** 207–211. ISSN: 14764687 (2015).

22. Bhatt, S. *et al.* Coverage and system efficiencies of insecticide-treated nets in Africa from 2000 to 2017. *eLife* **4,** e09672. ISSN: 2050-084X. http://elifesciences.org/content/early/2015/12/29/eLife.09672.abstract (2015).

23. Bertozzi-Villa, A. *et al.* Maps and metrics of insecticide-treated net access, use, and nets-per-capita in Africa from 2000-2020. *Nature Communications 2021 12:1* **12,** 1–12. ISSN: 2041-1723. https://www.nature.com/articles/s41467-021-23707-7 (2021).

24. Rathmes, G. *et al.* Global estimation of anti-malarial drug effectiveness for the treatment of uncomplicated Plasmodium falciparum malaria 1991-2019. *Malaria journal* **19.** ISSN: 1475-2875. https://pubmed.ncbi.nlm.nih.gov/33081784/ (2020).

25. Weiss, D. J. *et al.* Global maps of travel time to healthcare facilities. *Nature Medicine,* 1–4. ISSN: 1546170X (2020).

26. Ton, J. F., Flaxman, S., Sejdinovic, D. & Bhatt, S. Spatial mapping with Gaussian processes and nonstationary Fourier features. *Spatial Statistics* **28,** 59–78. ISSN: 2211-6753 (2018).

27. Milton, P., Coupland, H., Giorgi, E. & Bhatt, S. Spatial analysis made easy with linear regression and kernels. *Epidemics* **29.** ISSN: 1878-0067. https://pubmed.ncbi.nlm.nih.gov/31561884/ (2019).

28. Griffin, J. T. *et al.* Reducing Plasmodium falciparum Malaria Transmission in Africa: A Model-Based Evaluation of Intervention Strategies. *PLOS Medicine* **7,** e1000324. ISSN: 1549-1676. https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000324 (2010).

29. Sonnewald, M., Dutkiewicz, S., Hill, C. & Forget, G. Elucidating ecological complexity: Unsupervised learning determines global marine eco-provinces. *Science advances* **6,** eaay4740. ISSN: 23752548. http://advances.sciencemag.org/ (2020).

30. Sirovich, L. & Kirby, M. Low-dimensional procedure for the characterization of human faces. *JOSA A, Vol. 4, Issue 3, pp. 519-524* **4,** 519–524. ISSN: 1520-8532. https://www.osapublishing.org/viewmedia.cfm?uri=josaa-4-3-519%7B%5C&%7Dseq=0%7B%5C&%7Dhtml=true%20https://www.osapublishing.org/abstract.cfm?uri=josaa-4-3-519%20https://www.osapublishing.org/josaa/abstract.cfm?uri=josaa-4-3-519 (1987).

31. Proctor, J., Brunton, S., Brunton, B. & Kutz, J. Exploiting sparsity and equation-free architectures in complex systems. *The European Physical Journal Special Topics* **223,** 2665–2684. ISSN: 1951-6355. https://epjst.epj.org/articles/epjst/abs/2014/14/epjst22313002/epjst22313002.html (2014).

32. Eckart, C. & Young, G. The approximation of one matrix by another of lower rank. *Psychometrika 1936 1:3* **1,** 211–218. ISSN: 1860-0980. https://link.springer.com/article/10.1007/BF02288367 (1936).

33. Alter, O., Brown, P. O. & Botstein, D. Singular value decomposition for genome-wide expression data processing and modeling. *Proceedings of the National Academy of Sciences of the United States of America* **97,** 10101. /pmc/articles/PMC27718/%20/pmc/articles/PMC27718/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC27718/ (2000).

34. Xu, D. & Tian, Y. A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science 2015 2:2* **2,** 165–193. ISSN: 2198-5812. https://link.springer.com/article/10.1007/s40745-015-0040-1 (2015).

35. Rodriguez, M. Z. *et al.* Clustering algorithms: A comparative approach. *PLOS ONE* **14,** e0210236. ISSN: 1932-6203. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0210236 (2019).

36. Swarndepp, S. J. & Pandya, S. An Overview of Partitioning Algorithms in Clustering Techniques. *International Journal of Advanced Research in Computer Engineering & Technology* **5,** 1943–1946. ISSN: 1098-6596. arXiv: arXiv:1011.1669v3. http://itd.idaho.gov/ohs/2011Data/Analysis2011.pdf (2016).

37. MacQueen, J. Some methods for classification and analysis of multivariate observations. *https://doi.org/* **5.1,** 281–298. https://projecteuclid.org/ebooks/berkeley-symposium-on-mathematical-statistics-and-probability/Proceedings-of-the-Fifth-Berkeley-Symposium-on-Mathematical-Statistics-and/chapter/Some-methods-for-classification-and-analysis-of-multivariate-observations/bsmsp/1200512992 (Jan. 1967).

38. Smith, T. *et al.* Mathematical modeling of the impact of malaria vaccines on the clinical epidemiology and natural history of Plasmodium falciparum malaria: Overview. *The American journal of tropical medicine and hygiene* **75,** 1–10. ISSN: 0002-9637. https://pubmed.ncbi.nlm.nih.gov/16931810/ (2006).

39. Kiware, S. S. *et al.* Attacking the mosquito on multiple fronts: Insights from the Vector Control Optimization Model (VCOM) for malaria elimination. *PLOS ONE* **12,** e0187680. ISSN: 1932-6203. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0187680 (2017).

40. Penny, M. A. *et al.* Public health impact and cost-effectiveness of the RTS,S/AS01 malaria vaccine: a systematic comparison of predictions from four mathematical models. *The Lancet* **6736,** 1–9. ISSN: 01406736. http://linkinghub.elsevier.com/retrieve/pii/S0140673615007254 (2015).

41. Guelbéogo, W. M. *et al.* Variation in natural exposure to anopheles mosquitoes and its effects on malaria transmission. *eLife.* ISSN: 2050084X (2018).

42. Vaisala. *HUMIDITY CONVERSION FORMULAS Calculation formulas for humidity* tech. rep. (2013). www.vaisala.com.

43. Institute for Disease Modeling. *Documentation — IDMOD* https://www.idmod.org/documentation (2021).

44. McCarthy, K. A., Wenger, E. A., Huynh, G. H. & Eckhoff, P. A. Calibration of an intrahost malaria model and parameter ensemble evaluation of a pre-erythrocytic vaccine. *Malaria Journal* **14,** 6. http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed%7B%5C&%7Did=25563798%7B%5C&%7Dretmode=ref%7B%5C&%7Dcmd=prlinks%20papers3://publication/doi/10.1186/1475-2875-14-6 (2015).

45. World Health Organization and RBM Partnership to End Malaria. *WHO — High burden to high impact: a targeted malaria response* tech. rep. (2018).

46. WHO. *World Malaria Report 2020* tech. rep. December (2020), 238.

47. Beck, H. E. *et al.* Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data* **5.** /pmc/articles/PMC6207062/%20/pmc/articles/PMC6207062/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6207062/ (2018).

Chapter 2

Projecting Eradication Feasibility Using Archetypes

> But the accrual of those four straight lines crossed by a fifth became disheartening. Each neat bundle was meant to organize time and bolster the belief that it was finite and that when some was used up, there would be less of it to come. As if time could be ticked off and then a specific future would be reached...
>
> Marisa Silver
> *Little Nothing*

2.1 Introduction

Chapter 1 described in detail the rationale, methodology, and considerations of an archetypes-based approach to malaria intervention impact assessment. The present chapter provides more context and history for the circumstances under which this framework was developed, describes the quantitative results of an early archetypes-based exercise, and discusses common tradeoffs encountered when models are commissioned to inform real-world policy.

The global eradication of malaria has been a goal of "tropical health" almost since the founding of the field, with the world's first large-scale eradication campaign attempted in the 1950s [1]. This effort led to malaria elimination in 15 countries, but did not come close to eradication. After the failure of this campaign, international energy to eradicate malaria was low until Bill and Melinda Gates' bombshell 2007 announcement that malaria eradication would be one of their foundation's largest goals [2]. The subsequent funding and lobbying from the Bill and Melinda Gates Foundation (BMGF) has transformed the landscape of malaria policy globally, re-centering malaria as a top priority in many countries where it is endemic. However, malaria control and elimination has grown no less complex since 1969, with perennial challenges of drug resistance, insecticide resistance, diagnostic tool evasion, subpatent infection detection, and the sheer volume of malaria cases and deaths experienced by so many across the globe.

While eradication was and remains BMGF's stated goal for malaria, and a number of World Health Organization (WHO)-backed consortia have mapped strategic pathways to eradication [3], no formal timeline to eradication had been posited or maintained prior to 2018. In this year, two independent groups with overlapping membership were simultaneously exploring the risks and benefits of setting a target date for eradication. The first, the WHO Strategic Advisory Group for Malaria Eradication (SAG_{ME}), was a working group of one of the world's largest international policy organizations led by WHO Global Malaria Programme director Dr. Pedro Alonso [4]. The other, the Lancet Commission for Malaria Eradication (LCME), was a privately-funded project led by Sir Richard Feachem and based out of the University of California San Francisco [5]. Both groups commissioned a range of epidemiologic, economic, and climate-related analyses exploring different facets of the challenge facing eradication efforts. The Malaria Atlas Project (MAP) was asked by both groups to generate a range of malaria maps under different environmental and malaria-control scenarios. This chapter describes the results of that exercise,

conducted in collaboration with the Institute for Disease Modeling (IDM).

MAP and IDM combined geospatial and mechanistic modeling strategies to conduct this analysis. First, pre-learned geospatial relationships were used to understand what impact non malaria-control secular trends such as climate change, land use change, and demographic shifts might have on malaria in the absence of changes in malaria policy. Next, an archetypes-based mechanistic model framework was developed to test different malaria intervention scenarios and determine what mix of current and future tools might foster eradication.

Results from this analysis were presented to both groups, generating a lively debate which resulted in two different policy stances: the SAG_{ME} determined that it is premature to set a target date for malaria eradication, while the LCME concluded that eradication within the next 30 years was "ambitious, achievable, and necessary" [5]. Drivers and implications of these divergent decisions are highlighted in the discussion section.

This analysis demonstrates a potential use case for archetypes-based approaches to modeling, highlights both strengths and weaknesses of disease modeling in highly speculative future paradigms, and brings to the forefront questions and considerations about the relationship between modeling and decision making in the international policy context.

2.2 Methods

The goal of this analysis was to assess: Under optimistic malaria control circumstances, is eradication feasible by 2050 with currently-available tools? If not, what type of new technology is most needed? "Optimistic", operationally, meant that this analysis was run in scenarios with no drug or insecticide resistance, no importation

of cases from one location to another, and no disruptive shocks to health systems. Furthermore, interventions were simulated at a much larger scale than has historically been the case. For example, insecticide-treated net scenarios distributed nets once a year, instead of once every three years as is standard in Africa, and indoor residual spraying campaigns were simulated to cover all malaria-endemic locations, while in reality these interventions are geographically focal. A costing analysis was not conducted by this group, though collaborators conducted an independent economic exercise [5].

When considering long-term disease trends, it is important to consider those factors exogenous to disease control that may still impact burden. Even if malaria control does not change, the world will look very different in 2050– climatically, economically, demographically, and environmentally. MAP developed a geospatial modeling framework to learn the relationship between these factors and malaria burden for the past and present of the disease, and these learned relationships were used to generate projected future scenarios in which intervention coverage stays at 2016 levels but other secular trends change.

Having taken into account the effect of secular trends, the next step was to estimate how malaria burden would be impacted by changes to malaria control policies. For this step, the EMOD mechanistic model was used in an archetypes-based framework to generate impact estimates for 152 different intervention packages, covering a wide range of existing vector control and case management policies as well as novel interventions such as Attractive Targeted Sugar Bait (ATSB)s, pre-erythrocytic and transmission-blocking vaccines (PEVs and TBVs), and monoclonal antibodies (mAbs).

The analysis was conducted globally, but this chapter will focus on the African setting because elimination was achieved easily in South American and Southeast Asian environments under the intervention scenarios described, and because these

lower-burden settings were run using a heavily simplified archetypes framework with a single archetype covering each continent.

2.2.1 Burden Change Due to Secular Trends

The work in this subsection was led by Drs Samir Bhatt and Peter Gething, but is described here as it forms a crucial set of inputs to the novel archetypes analysis that is the main focus of this chapter. Writing is by Amelia Bertozzi-Villa, with contributions from the original analysts.

For this analysis, MAP generated high-resolution estimates of *Plasmodium falciparum* parasite rate among 2-10 year olds ($PfPR_{2\text{-}10}$) using a machine learning approach that has been described in detail elsewhere [6, 7]. Briefly: a single-layer Rectified Linear Unit neural network with 10,000 nodes was used to learn complex nonlinear relationships between (a) malariometric data including more than 50,000 independent observations of $PfPR_{2\text{-}10}$ at defined point locations worldwide; (b) data on Insecticide-Treated Net (ITN), Indoor Residual Spraying (IRS), and Artemether-Lumefantrine Case Management (AL CM) coverage at point locations and across national and subnational regions; and (c) gridded geospatial environmental covariates characterizing the natural and human micro-environments.

To generate future-projected maps of PR in 2050, the historical relationships learned in the neural network analysis were used to predict $PfPR_{2\text{-}10}$ for a set of projected environmental covariates (Table 2.1). Whenever possible, publicly-available products designed by field-specific experts were utilized for these projected values. In particular, heavy use was made of the projected rasters associated with the Shared Socioeconomic Pathways/Representative Concentration Pathways (SSP/RCP) scenarios established by the Intergovernmental Panel on Climate Change [8]. From the five SSP and four commonly-used RCP scenarios available, the SSP2 “Middle of

the Road" and RCP "CAM-RCP4.5" scenarios were selected, reflecting expectations of stable social, economic and technological trends and a greenhouse gas emission radiative forcing target of 4.5 W/m^2 after 2100. The SSP5 "Conventional Development" scenario was also tested, and had a negligible impact on results. Three covariate surfaces used in the PR model (Night-time light brightness, impervious surfaces, and improved housing) had no externally available projections. For these, rasters from the latest available year were projected forward via a national-level linear regression on the projected gross domestic product and urbanicity covariates.

Two 2050 $PfPR_{2\text{-}10}$ transmission surfaces were generated: one using 2016 AL CM, IRS, and ITN intervention surfaces, and one with no malaria interventions. The former was used to show the potential impact of secular trends alone on malaria transmission, while the second was used as a baseline for the EMOD simulation results described below.

Covariate	Source
Night-time light brightness	MAP Regression
% houses electrified	MAP Regression
% housing with modern construction	MAP Regression
% impervious surfaces	MAP Regression
Population density	Climate and Global Dynamics Laboratory, National Center for Atmospheric Research
Gross domestic product (pixel level)	Land-Use Harmonization 2 project
% urban population	Climate and Global Dynamics Laboratory, National Center for Atmospheric Research
Land surface temperature (max)	Consultative Group for International Agricultural Research
Land surface temperature (min)	Consultative Group for International Agricultural Research
Land surface temperature (seasonality)	Consultative Group for International Agricultural Research
Land surface temperature (diurnal range)	Consultative Group for International Agricultural Research
Precipitation (max)	Consultative Group for International Agricultural Research
Precipitation (min)	Consultative Group for International Agricultural Research
Precipitation (seasonality)	Consultative Group for International Agricultural Research
Precipitation (diurnal range)	Consultative Group for International Agricultural Research
% forested primary land	Land-Use Harmonization 2 project
% non-forested primary land	Land-Use Harmonization 2 project
% managed pasture	Land-Use Harmonization 2 project
Wood harvest area from primary forest (i.e. deforestation)	Land-Use Harmonization 2 project
% irrigated	Land-Use Harmonization 2 project

Table 2.1: Projected environmental covariate sources.

2.2.2 Additional Burden Reduction Under Intervention Scenarios

To simulate the effect of different malaria interventions on this new 2050 prevalence map, mechanistic disease models were used. Chapter 1 goes into detail on the

archetypes framework that was developed expressly for this exercise, including a range of different choices available at different steps. The choices made for this early version of the analysis are as follows.

Clustering covariates were selected from the MAP database, and included rainfall from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) project [9], a Temperature Suitability Index (TSI) metric which reflects the temperature-based ability of mosquitoes to reproduce and survive [10], and relative mosquito abundances of *Anopheles arabiensis, funestus*, and *gambiae* in Africa [11]. Relative vector abundance is a static metric, while rainfall and TSI are time-varying and monthly. Synoptic values of time-varying estimates were calculated from 2000 to 2016 to generate twelve-monthly metrics showing mean seasonal changes. Therefore, the final covariate stack was 27 dimensions (12 layers each of rainfall and TSI, plus three mosquito layers). Covariates were not masked to *Plasmodium falciparum* transmission limits before clustering, and were not rescaled.

Singular Value Decomposition (SVD) was conducted for each continent using the `svd` command in `R` version 3.6.2. Enough singular values were included for further analysis to cover at least 95% of the variation in the original dataset.

K-means clustering was performed on a range of clusters from three to 15. K-means was applied using the `kmeans` command in the `stats` R package (version 4.0.5). A cluster count of six was selected heuristically as a value that displayed a range of meaningfully different seasonal climatic trends without excluding important seasonality types or including redundant seasonalities.

Representative site selection was limited by EMOD's climate capabilities at the time. When this analysis was conducted, use of ERA5 data as climate input files into EMOD had not yet been developed. Instead, climate files were generated via IDM's in-house climate model, based on weather station data [12], but this climate

modeling was only available in certain countries. As a result, representative sites were selected manually for each archetype based on climate data availability.

EMOD simulations were run identically to Chapter 1, with two differences: Country-specific birth rates for each representative site were used instead of a single uniform birth rate, and in-house climate files were used, with a 15-year no-intervention simulation to establish population immunity followed by a 3-year intervention period. See Table 1.1 for species-specific mosquito behavior parameters, Table 2.2 for site-specific mosquito behavior parameters, and Table 1.2 for intervention descriptions used in the model.

After running simulations, the intervention-deleted version of the 2050 prevalence map was used as a baseline transmission surface, to account for the fact that simulations were run without historical interventions.

2.2.3 Burden Estimation

The incidence associated with each counterfactual $PfPR_{2\text{-}10}$ maps was calculated using an established prevalence-to-incidence conversion [13], and population-at-risk determined by summing the population projected to reside in every pixel according to the population covariate used in the secular trends analysis.

To assess whether policy recommendations would have been different under the more rigorous archetypes framework described in Chapter 1, intervention impact results from that analysis were applied to the intervention-deleted 2050 map to compare to the present results. See Table 2.3 for a summary of differences between the two methods.

Site #	**Description**	**Site Country**	***arabiensis* Proportion**	***funestus* Proportion**	***gambiae* Proportion**	**Anthrophily (%)**	**Endophily (%)**	**Total Indoor Biting (%)**
1	Lime, Coastal west Africa and Republic of the Congo	Nigeria	0.14	0.40	0.46	74	80	59
2	Teal, Arid northern and southern Africa	Ethiopia	1	0	0	65	50	32
3	Purple, Gabon, CAR, DRC, and Uganda	DRC	0	0.001	0.999	85	85	72
4	Gold, Sahelian	Burkina Faso	0.84	0.16	0	65	56	36
5	Blue, Southern Africa and central Tanzania	Mozambique	0.95	0.05	0	65	52	34
6	Magenta, Zambia, Malawi, north Mozambique, and Madagascar	Zambia	0.13	0.23	0.64	78	80	62

Table 2.2: EMOD Site Descriptions. CAR: Central African Republic; DRC: Democratic Republic of the Congo.

Analytical Step	Megatrends Version	New Version
Method of site selection within clusters	Ad hoc	Pixel closest to centroid center
Method of cluster count selection	Ad hoc	Elbow plot
Archetype climate source	CHIRPS rainfall and Temperature Suitability (both from MAP)	ERA5 Air Temp/Precipitation
Archetype: data layers used	Synoptic Rainfall Synoptic TSI Relative Vector Abundance	Synoptic Precipitation Synoptic Temperature Relative Vector Abundance
Is data rescaled before SVD?	No	Yes
Simulations climate source	IDM in-house climate files (from weather stations)	ERA5
Immune burn-in duration	15 years	39 years
Birth rate reference	World Bank: Individual countries	World Bank: All of sub-Saharan Africa

Table 2.3: Difference between old and new method.

2.3 Results

Synoptic temperature suitability index values ranged from 0 to 0.94, with a median value of 0.55 and an Interquartile Range (IQR) from 0.28 to 0.69. CHIRPS rainfall ranged from 0 to 1,082 mm/month, with a median value of 16.7 and an IQR from 1.9 to 105.7. Relative mosquito abundance values ranged from 0 to 1, with continent-wide proportions of 43.5% *arabiensis*, 36.3% *gambiae*, and 20.2% *funestus* among those pixels with mosquitoes. Figure 2.1 shows covariate distributions. Because covariates were not rescaled, the rainfall variable dominates both the distribution plots and the analysis, with a strong peak near zero and a smaller one near 150 mm/month.

Figure 2.2 shows the proportion of variance explained by the first five singular vectors in the SVD procedure. The first three singular vectors accounted for 96.1% of the variation and were retained for k-means. Figure 2.3 shows the elbow plot for the k-means procedure, which was not used to determine an optimal k in the

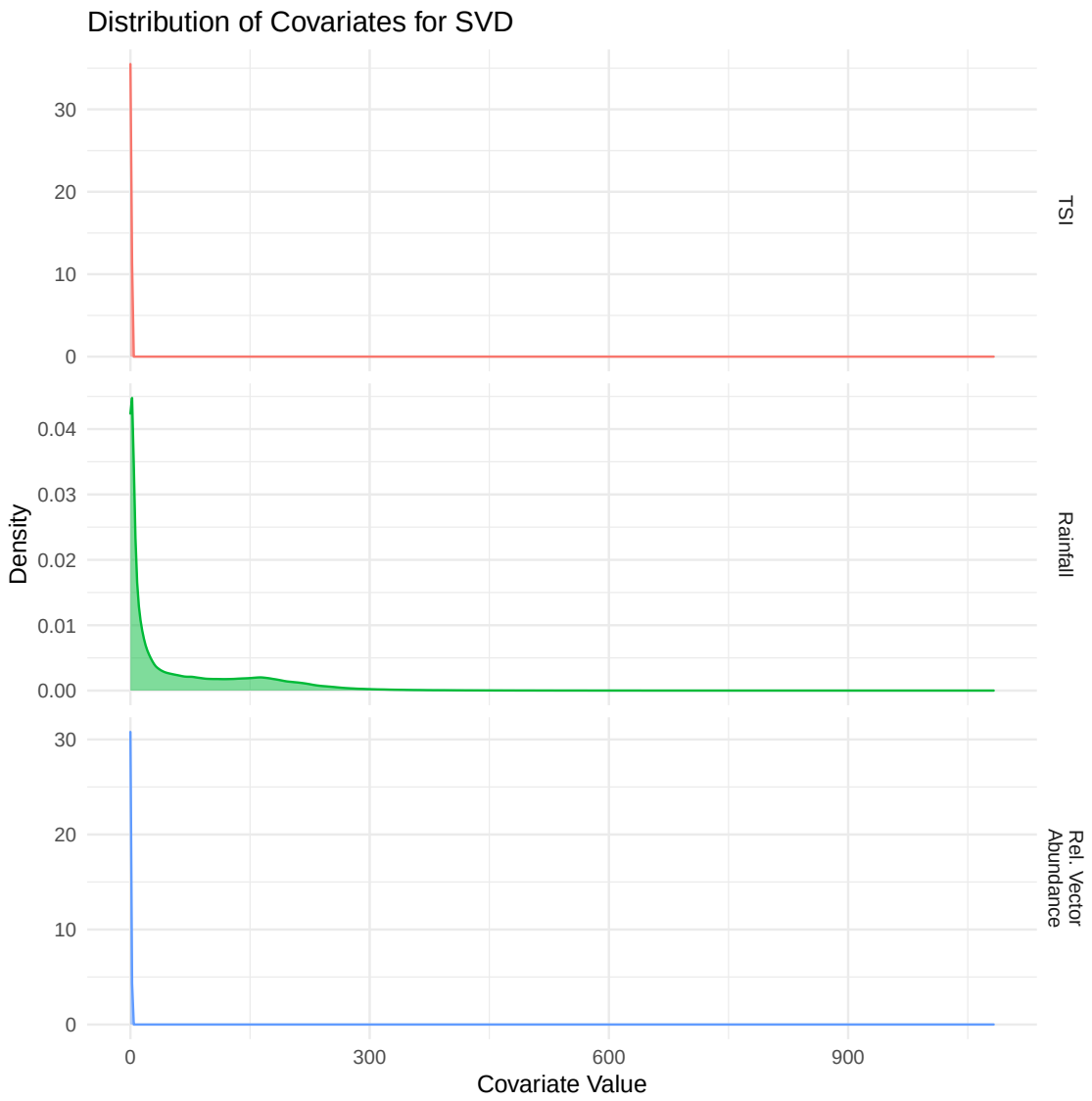


Figure 2.1: Distribution of covariates for Singular Value Decomposition (SVD) across all pixels in the analysis. Synoptic monthly Temperature Suitability Index (TSI) is a modeled covariate from MAP describing how amenable a location is to mosquito survival based on temperature. Synoptic monthly rainfall is from the CHIRPS project. Relative vector abundance, a static metric, tracks the proportion of *Anopheles arabiensis*, *funestus*, and *gambiae* in each pixel. Rainfall is in units of millimeters per month, while TSI and relative vector abundance are proportions ranging from zero to one. Because covariates were not rescaled, the rainfall distribution dominates both this plot and the SVD analysis.

original assessment. It shows that, while a k of six captures over 80% of total data variance as between-cluster variance, a slightly higher cluster count would have covered close to 90% of this ratio. Figure 2.4 shows a snapshot of the reprojected data points in 3D space, colored according to the k-means results. Black dots represent true cluster centroids, while gray dots represent the manually-selected representative sites. The shape of this data shows a centering around the origin which then flares out in three orthogonal wings. This structure is likely due to the un-standardized data inputs, which result in less intuitive cluster boundaries than those seen in the analogous plots in Chapter 1. Figure 2.5 shows archetype maps for all values tested. In broad strokes, the progression of archetypes in this setting tracks closely to that seen in Chapter 1. First, the three-archetype setting subdivides the continent into arid northern and southern regions, a seasonal band across coastal west Africa and the Sahel, and a less seasonal band from Gabon to Madagascar. As archetype counts increase, additional latitudinal bands appear in conjunction with the subdivision of central Africa along a southwestern-to-northeastern boundary cutting through southern Angola and Democratic Republic of the Congo (DRC), and isolation of specific regions in the Great Rift Valley and coastal west Africa. Despite these similarities, the maps in Figure 1.5 are notably more heterogeneous and less contiguous than those in Figure 2.5, especially in Ethiopia and around Lake Victoria. In addition, the inclusion of highly arid Saharan pixels in Figure 2.5 results in the continued grouping together of northern and southern deserts into a single archetype, whereas in Chapter 1 these two regions disaggregate as early as an archetype count of four.

Figure 2.6 shows the map and time series for the six-archetype setting used for the modeling analysis. In the time series, solid colored lines represent the median across the archetype, shaded areas indicate the interquartile range, and colored dotted lines indicate the 95% variance interval. Solid black lines represent the procedurally-selected climate values of the representative site for each archetype,

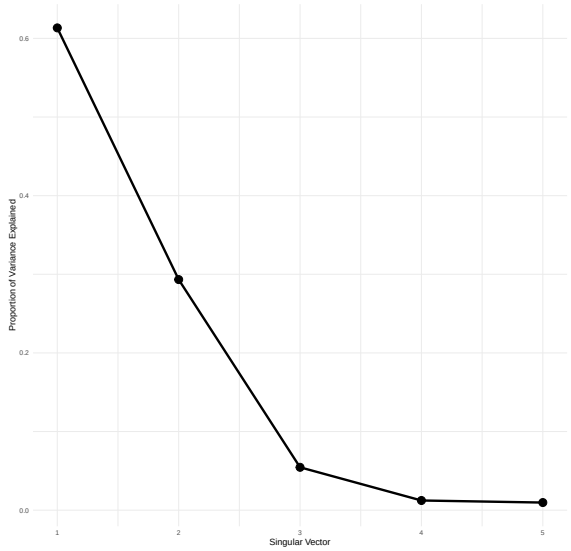


Figure 2.2: Proportion of variance explained by the first five singular vectors after Singular Value Decomposition (SVD). The first three vectors were retained for clustering analysis.

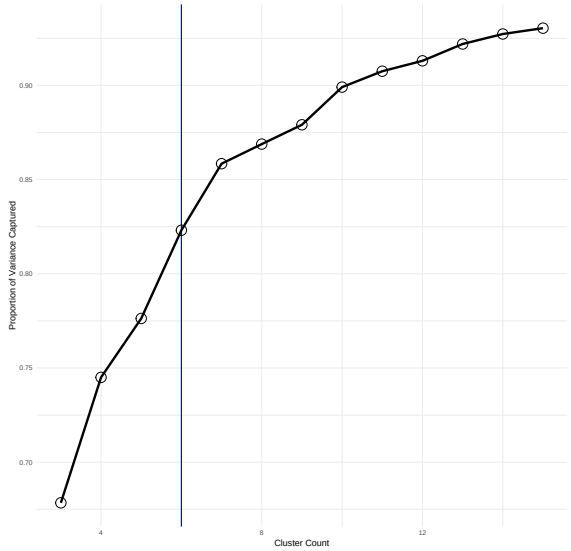


Figure 2.3: K-means elbow plot for k between 3 and 15. The x-axis shows cluster count, while the y-axis shows what proportion of total data variance is captured by between-cluster variance in each setting. A cluster count of six (vertical blue line) was chosen for future analysis based on visual examination of maps, not based on the amount of variance captured.



Figure 2.4: Top and side view of pixel values reprojected onto the first three singular vectors, after k-means clustering with a k of six. Colors refer to different clusters and match subsequent six-cluster plots. Black dots indicate true cluster centroids, while dark grey dots show manually-selected representative sites.

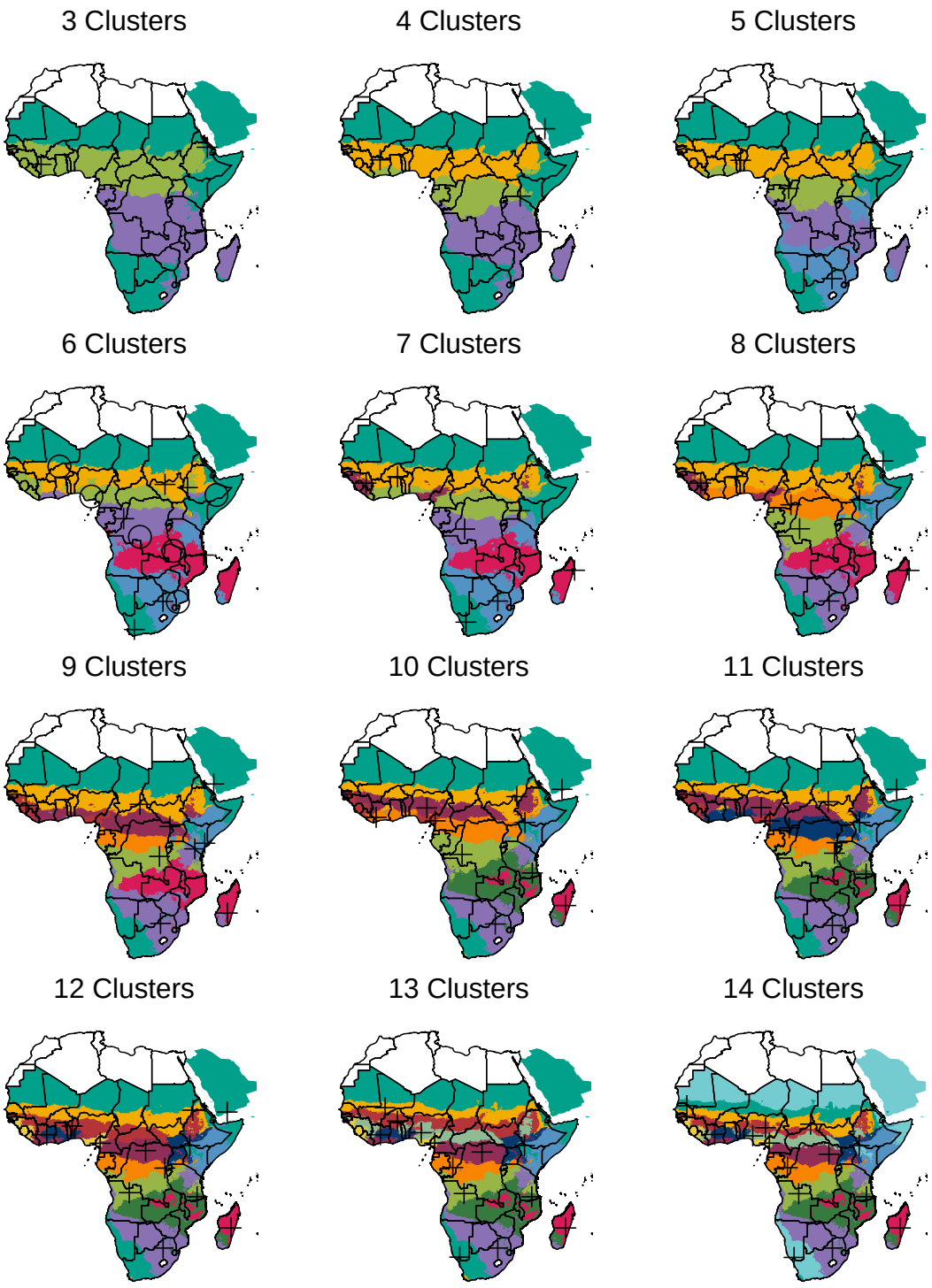


Figure 2.5: Cluster maps for k of 3 to 14. Black crosses indicate procedurally-generated representative sites (the pixels closest to cluster centroids), while the open black circles in the six-cluster plot show manually-selected sites.

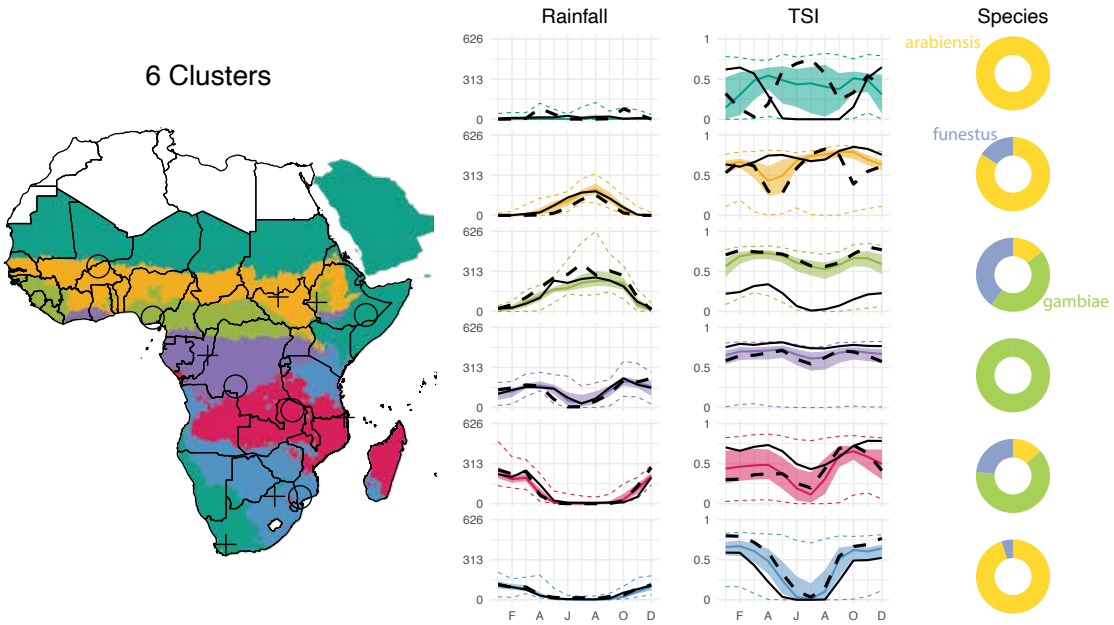


Figure 2.6: Maps and time series for the six-site setting. In the time series, solid colored lines represent the median across the archetype, shaded areas indicate the interquartile range, and colored dotted lines indicate the 95% variance interval. Solid black lines represent the procedurally-selected climate values of the representative site for each archetype, also indicated as black crosses on the map, while dashed black lines and circles on the map represent the manually-selected sites that were used in the modeling analysis. Doughnut plots show the relative vector abundance of the manually-selected representative sites.

also indicated as black crosses on the map, while dashed black lines and circles on the map represent the manually-selected sites that were used in the modeling analysis. Despite the fact that they were selected without a comparison to the archetype time series, temperature suitability index values in the manually-selected sites consistently match archetype medians more closely than the procedurally-selected sites. Rainfall values hew consistently close to archetype medians in both cases.

The six-archetype map is divided into largely latitudinal bands. The first (teal), covering the Sahara, the horn of Africa, and southern African deserts, features almost no rainfall, low temperature suitability, and an *arabiensis*-dominated landscape. Next are two landscapes with single-peaked seasonality: a somewhat seasonal, *arabiensis*-dominated Sahelian band stretching from Senegal to western Ethiopia

(gold), followed by an even more strongly seasonal archetype with a mix of all three vectors covering the far western coast, a band from southern Nigeria across the north edge of DRC, and a small section of western Ethiopia (lime). Moving south is a spatially contiguous, seasonally bimodal central African archetype stretching from Gabon through central DRC and into Uganda and featuring exclusively *gambiae* (purple). The fifth archetype (magenta) reaches from central Angola across Zambia, southern DRC, Malawi, northern Mozambique, and Madagascar, and is characterized by strongly seasonal rainfall and a *gambiae*-dominated vector landscape with some *funestus* and a small amount of *arabiensis*. Finally, a *funestus*-dominated archetype with less-seasonal rainfall, and highly seasonal temperature suitability covers three noncontiguous regions: north-central Tanzania, southern Angola through southern Mozambique and eastern South Africa, and the southern tip of Madagascar (blue). Section 2.5 shows similar maps for all cluster counts tested.

Figure 2.7 shows projected malaria prevalence in Africa from 2016 to 2050 if interventions were kept at 2016 coverage levels. All changes in this plot are due to projected changes in secular trends alone. Compared to 2016 levels, a substantial decline in transmission intensity is projected across the continent, with residual pockets of higher transmission in northern Mozambique, southern DRC, and scattered portions of west Africa. The black lines of Figure 2.8 translate this prevalence into incidence rates and absolute case counts. The incidence rate continentally is projected to fall from over 200 cases per 1,000 people per year to less than 150 by 2050. However, due to concurrent increases in population, absolute case counts are expected to stay almost constant at 200 million cases per year. Figure 2.8 also shows analogous incidence curves for some interventions highlighted in Figure 2.11, showing that a dramatic increase in ITN, IRS, and antimalarial coverage may bring the continent close to eradication, but that new tools may not add much additional benefit.

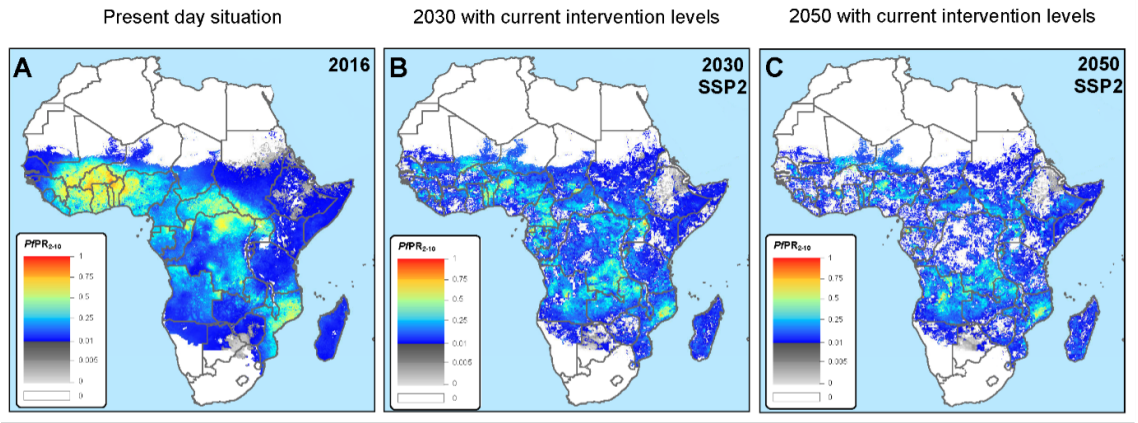


Figure 2.7: Projected malaria prevalence due to secular trends alone, holding interventions at 2016 coverage levels. The baseline 2016 map is shown in A, while B and C show projected maps for 2030 and 2050, respectively. Figure created by Peter Gething and Jennifer Rozier and used with permission.

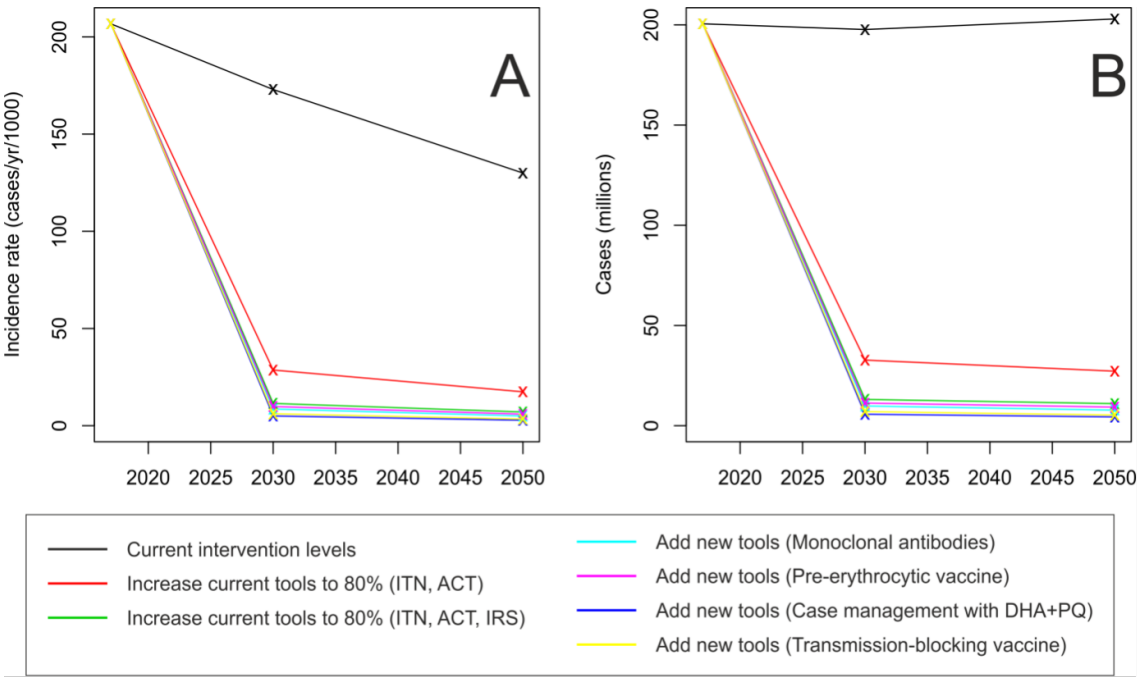


Figure 2.8: Selected intervention package results in terms of incidence rates (A) and total cases (B). Figure created by Peter Gething and Jennifer Rozier and used with permission. The label ACT here is analogous to "AL CM" in the remainder of the document, indicating Artemether-Lumefantrine Case Management.

Figure 2.9 shows intervention impact results for the same four interventions as showcased in Figure 1.7, for comparison. The first scenario, upper left map and dot-dashed lines, does not have interventions and therefore recreates the baseline prevalence map, which in this case represents a 2050 scenario in which malaria interventions have never been deployed, with reductions in transmission intensity driven by secular trends alone. The second, upper right map and dotted lines, shows a complex intervention mix of annual ITN distributions with moderate coverage (40%), annual IRS campaigns with 20% coverage, and low access and use of AL CM (20% of clinical cases receiving treatment). The third, lower left map and solid lines, shows three years of high antimalarial access and use (80% of clinical cases receiving treatment), but no other interventions. The fourth, lower right map and dashed lines, shows a scenario in which the only intervention is a hypothetical ATSB with a 3% mortality rate. For more intervention details see Table 1.2. Starting as they do from a lower initial transmission intensity, all three intervention packages tend rapidly toward elimination or near-elimination settings. As in the analogous plot in Chapter 1, initial transmission intensity is the strongest driver of final transmission intensity, and the antimalarial-only scenario is more effective at higher transmission intensities and approximately as effective at lower transmission intensities compared to vector control interventions. Also similar to the setting in Chapter 1, mosquito-specific interventions show much more variation across sites than non-mosquito interventions, following intuitive patterns relating to indoor biting behavior (Figure 2.10). Maps and curves for all intervention packages are available under the "6-Site Setting" label at `https://institutefordiseasemodeling.github.io/archetypes-intervention-impact/`.

Figure 2.11 shows the intervention maps that were highlighted for the stakeholders of the eradication feasibility analysis. These intervention packages include high ITN and antimalarial coverage with no IRS; high coverage of ITNs, IRS, and antimalarials; mid-level coverage of mAbs, a TBV, or a PEV while maintaining

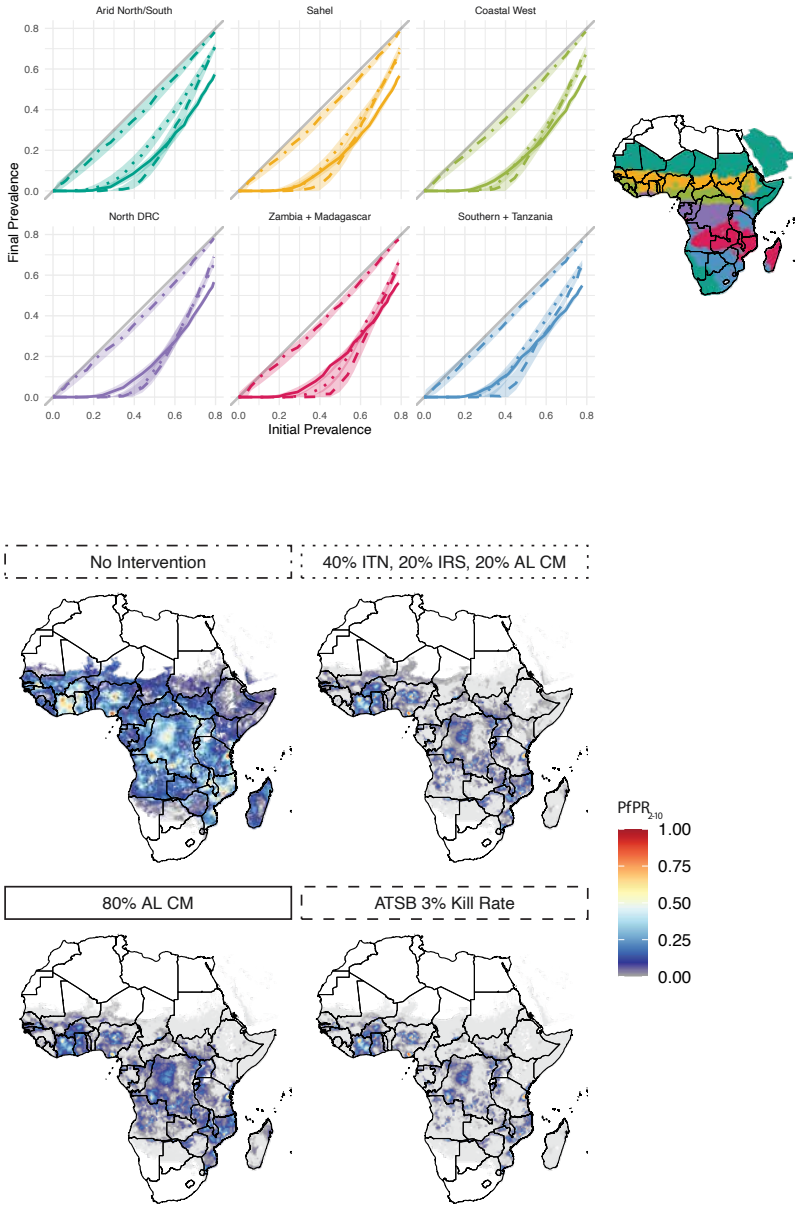


Figure 2.9: Example intervention impact results for four of the 152 interventions tested. The line plots (top) show intervention impact curves disaggregated by archetype, with archetypes colored according to the map (top right). The four prevalence maps (bottom) show hypothetical *Plasmodium falciparum* parasite rate among 2-10 year olds ($PfPR_{2\text{-}10}$) in 2050 for each scenario. These maps were created by starting from a baseline map of a counterfactual 2050 in which Africa never had malaria interventions but did experience the burden changes of secular trends. For each initial prevalence value in the baseline map, a new prevalence value was found using the lines in the top set of plots. The linetype of subplot borders in the lower plots matches the linetype of interventions in the upper plots. For more detail about interventions see Table 1.2.

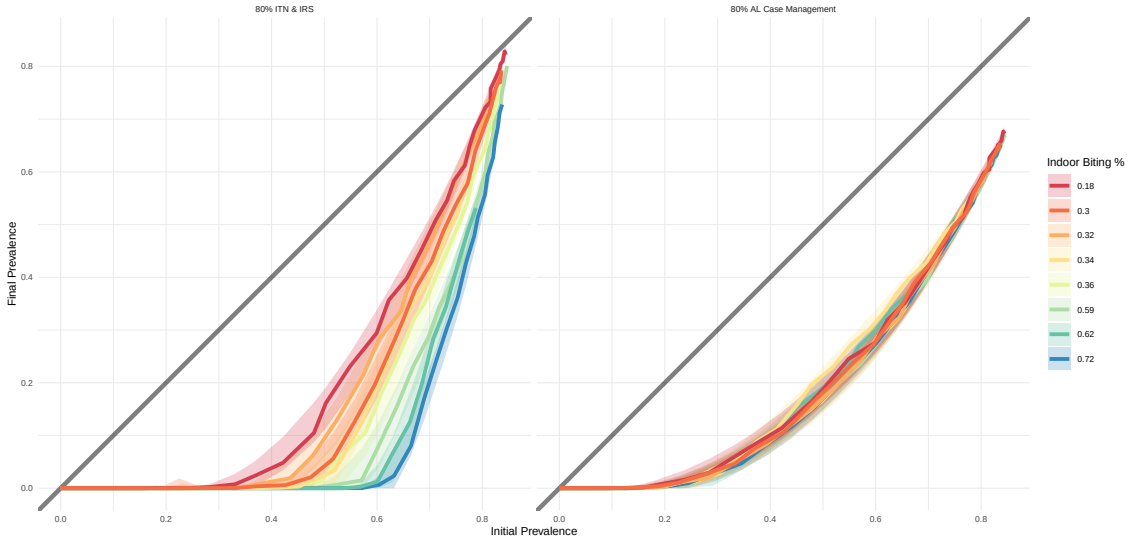


Figure 2.10: Sensitivity of sites to biting intensity for vector control vs non-vector-control interventions. In both plots, each line represents an archetype's representative site, colored according to its indoor biting percentage as determined by its mosquito species mix. The left panel shows an intervention setting with 80% coverage of Insecticide-Treated Net (ITN)s and Indoor Residual Spraying (IRS), both of which target mosquitoes indoors. This intervention package, as expected, is much more effective in areas with a higher percentage of bites occurring indoors. The right panel demonstrates the inverse property: under an intervention package that includes only anti-malarial drugs and therefore does not directly target mosquitoes at all, there is no differential impact across indoor biting intensities.

high background coverage of nets, drugs, and IRS; the replacement of artemisenin-lumefantrine antimalarials with longer-lasting dihydroartemisinin-piperaquine drugs (DP CM) while maintaining high ITN and IRS coverage; widespread deployment of moderately effective ATSBs in the absence of any other interventions; and widespread deployment of extremely effective ATSBs along with high coverage of nets, drugs, and IRS. Only this latter scenario succeeds in completely eradicating malaria.

To assess whether policy recommendations would have been different under the more rigorous archetypes framework described in Chapter 1, intervention impact results from that analysis were applied to the intervention-deleted 2050 map to compare to the present results. Figure 2.12 shows distributions of the difference between Population at Risk (PAR) in each of the two methods. Percent differences between the two methods cluster near zero, though the original six-cluster analysis predicts a somewhat larger population-at-risk in most scenarios. Using the six-archetype method as a baseline, there is a positive percent difference between methods in 99 cases, a negative percent difference in 50 cases, and 4 cases in which eradication was achieved in both methods. The mean and median percent difference in PAR between old and new methods was 8.7 million and 5.7 million people, respectively. Percent differences between old and new PAR ranged from -37.5% in the scenario with 40% coverage of ITNs, IRS, and antimalarials plus 15% kill rate of ATSBs, in which the old version predicts a PAR of 7.2 million and the new version a PAR of 9.9 million, to 100% in three cases where the new PAR is zero and the old was nonzero: a 25% kill rate ATSB with no other interventions, a 15% kill rate of ATSBs with 60% coverage of ITNs, IRS, and antimalarials, and 15% kill rate of ATSBs with 80% coverage of ITNs, IRS, and antimalarials. Absolute PAR differences ranged from -121.9 million in the scenario with 20% coverage of ITNs, IRS, and antimalarials, to 59.2 million in the scenario with 20% coverage of ITNs, no IRS, and 80% coverage of antimalarials. Table 2.4 gives intervention descriptions and PAR comparisons for all interventions. Maps comparing the impact of

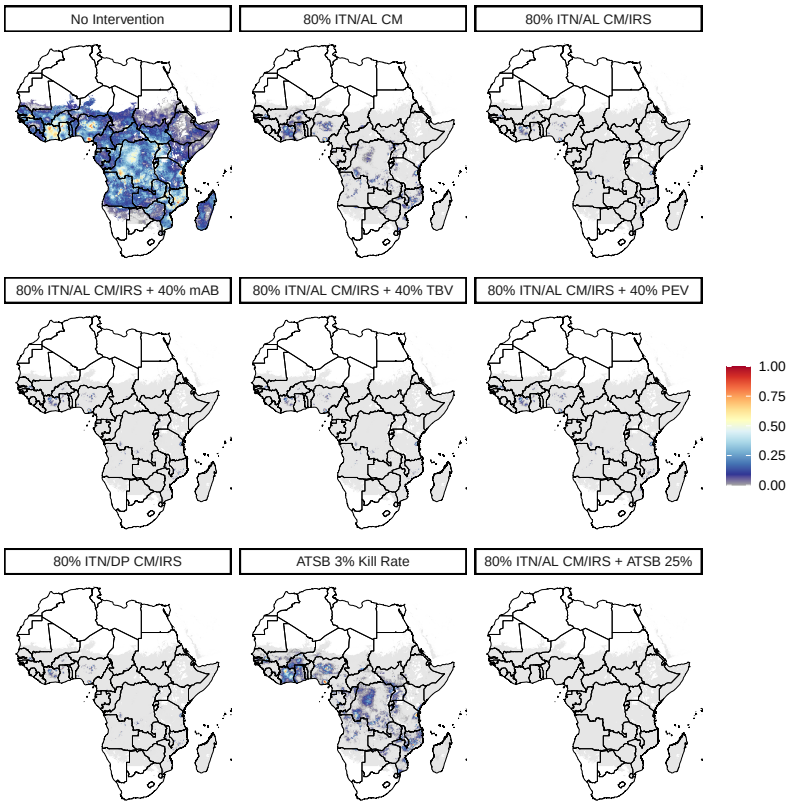


Figure 2.11: Intervention impact maps of $PfPR_{2\text{-}10}$, analogous to those in Figure 2.9, for the interventions presented to eradication project stakeholders. ITN, insecticide-treated net; AL CM, artemether-lumefantrine case management; IRS, indoor residual spraying; mAB, monoclonal antibodies; TBV, transmission-blocking vaccine; PEV, pre-erythrocytic vaccine; DP CM, dihydroartemisinin-piperaquine case management; ATSB, attractive targeted sugar bait.

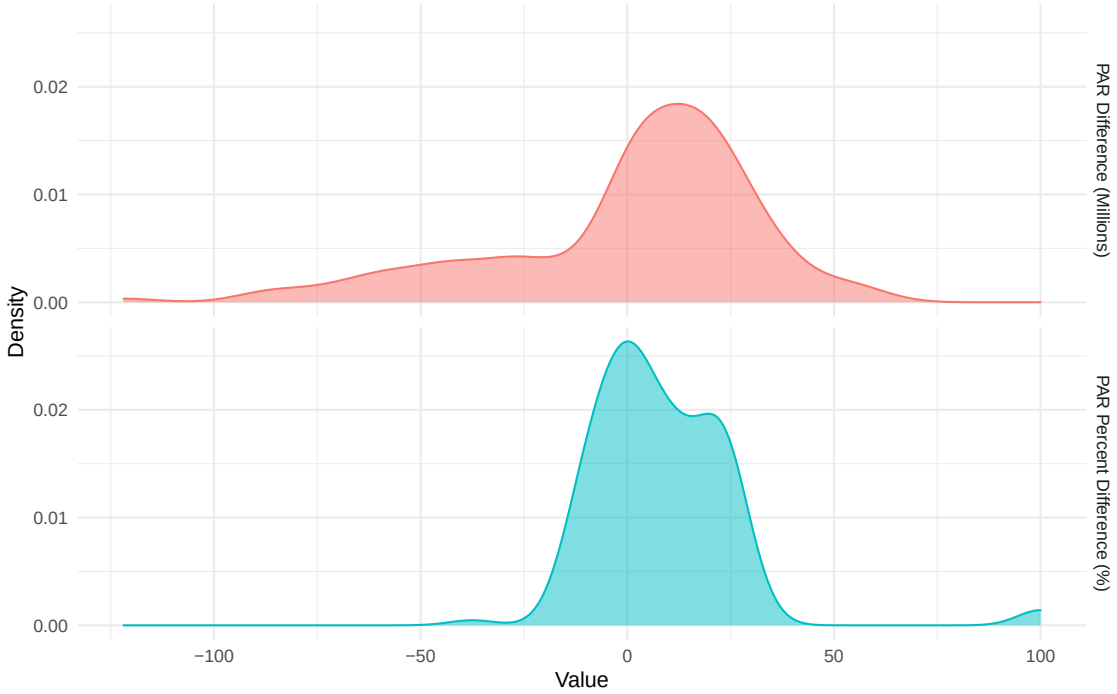


Figure 2.12: Comparison of results using the archetypes and sites from Chapter 1 compared to Chapter 2. For each of the 152 interventions, the difference in population-at-risk (PAR) was compared between the two methods. The distribution of absolute differences in PAR is shown in the top panel, with percent differences shown in the bottom panel. In all cases, the Chapter 2 is the reference value.

the two methods for all interventions are available under the "Methods Comparison" label at https://institutefordiseasemodeling.github.io/archetypes-intervention-impact/.

2.4 Discussion

This modeling analysis found that, even under highly optimistic climatic, economic, and demographic conditions, currently-available tools would be insufficient to achieve malaria eradication by 2050, though they would drastically reduce malaria burden and achieve elimination in many settings. Further, it found that novel interventions would need to be applied at unprecedented coverages and efficacies in order to attain eradication in this time frame. Further challenges to eradication that

were not included in the exercise include expanding drug and insecticide resistance, re-seeding of malaria due to human movement, and catastrophic structural effects of climate change (as opposed to the direct effects of climate change on mosquito larval habitat). From this single set of results, the two policy stakeholders came to very different conclusions. While neither went so far as to dismiss the goal of malaria eradication, the WHO SAG_{ME}'s report took the stance that the time had not yet come to craft an eradication timeline. Instead, they argued, policy makers at every scale should work to reduce malaria burden and build robust systems to prepare a future generation for eradication planning. The LCME, by contrast, concluded that all the stated challenges were solvable and released a report describing malaria by 2050 as "ambitious, achievable, and necessary".

While these divergent interpretations of the same analysis may at first seem surprising, they make sense in consideration of the roles and responsibilities of the two different groups. WHO and its satellites are the most broadly-recognized authorities on health best practices in the world. Policy decisions that affect millions of lives are made based on the recommendations printed on WHO letterhead. The group's words are powerful, and with this power comes the responsibility to exercise great care in determining what to write. It is reasonable that WHO would avoid setting eradication targets in the absence of strong evidence in the feasibility of such a proposal, especially given historic criticism regarding global WHO targets that remain unmet. The LCME was working in many ways under the opposite set of incentives. As a private group with no formal policy-making power, anything short of a bold and urgent call for eradication would be insufficient to grasp public attention. By adding energy to eradication efforts, the LCME could perhaps add momentum to national or local elimination efforts that had been stalling. For this group, setting even an unrealistic target has value because it recenters the discourse and, perhaps, promotes friendly country-to-country rivalry. However, this eagerness to set such bold targets required the group to claim a global capacity for malaria campaign ef-

ficacy and a trust in the power of emerging technology that was less well-supported by both historical evidence and the modeling exercise than the more conservative stance of the WHO SAG_{ME}.

The analysis presented here represents the successful deployment of an innovative, flexible, and customizable modeling strategy designed explicitly for collaboration with malaria policymakers. This modeling framework is sufficiently complex to capture important nuances of malaria as a multifaceted and spatially heterogeneous disease, while also simple enough to communicate intuitively to non-modeling audiences. While the two main stakeholders for this analysis came to different policy conclusions, both did so with a clear-headed interpretation of the modeling results that was only possible through strong bidirectional communication. This work also showcases the benefit of collaborations between modeling groups, allowing the strengths of different quantitative approaches to support each other. As described in Chapter 1, the archetypes-based modeling strategy itself is a novel and computationally effective approach to quickly and informatively assess the impact of potential malaria interventions even in low-data settings, and this framework has already proven valuable in a range of use cases beyond the eradication exercise described here.

This collaboration and model development process faced limitations. Because model development was by necessity linked to the deadlines of policy stakeholders, several modeling choices were made based on heuristics and convenience rather than on best practices for the method under consideration. In this instance, a *post hoc* analysis was conducted with a more rigorous set of archetype creation criteria, which was then compared to the original results. Fortunately, the two methods showed good agreement across all intervention strategies, but there was no guarantee that this would be the case. As policymakers and modelers work together more closely, both groups are becoming more familiar and comfortable with tradeoffs of rigor

and speed, and are increasingly able to communicate honestly about priorities and jointly determine an acceptable analytical path.

As quantitative methods for global health have become more mainstream, increasing scrutiny has been given to communication of the uncertainty accompanying modeling results to avoid false confidence in the estimates provided. In this particular instance, however, uncertainty was not considered because the research questions were so speculative that the notion of uncertainty propagation began to lose meaning. Almost by definition, any uncertainty bounds around the estimates presented here would span the entire range of possible values, which would make them entirely unhelpful. Or, if uncertainty bounds did show a narrower scope, these results might misleadingly suggest more confidence in the future than this analysis was able to provide. Instead, modelers chose to eschew formal uncertainty estimates and focus on clearly communicating that all the results shown were scenario-based speculations rather than formal predictions of the future.

Limitations of the archetypes strategy are described at length in Chapter 1. The approach used here to determine the impact of secular trends on malaria is based upon well-established statistical relationships, but makes several simplifying assumptions: first, that these relationships will remain unchanged as environmental, economic, and demographic patterns change over time; and second, that these statistical relationships are sufficient for future projections, even though they describe complex and dynamical systems with deeply entangled causal pathways. The first assumption, while imperfect, is the best choice available given a lack of knowledge about how these relationships will change in the future. The second assumption is more complex, and Nissan et al. [14] have written a detailed and nuanced rebuttal emphasizing that taking the nonlinear impacts of climate change into account will be instrumental in developing a robust and resilient malaria control strategy as the world moves deeper into the 21st century. Nissan et al.'s work also makes the crucial

observation that attempting to predict a single future with any degree of accuracy is a far less effective strategy than is predicting a wide range of possible futures and building global health infrastructures and social safety nets that can show flexibility and resilience in the face of this spectrum of unknown challenges. The archetypes modeling framework could be well-suited to such a purpose, and future work will center this philosophy.

Collectively, Chapter 1 and Chapter 2 have introduced a novel method for assessing large-scale intervention impact at arbitrarily small spatial resolutions. Nuances, strengths, and weaknesses of this format have been presented along with a detailed use case describing the utility of such a framework in real-world policy settings. This framework continues to grow, evolve, and find additional use cases, including but not limited to evaluating whether larger insecticide-treated net campaigns could mitigate the impact of COVID-related antimalarial stockouts, assessing potential burden reduction implications of ITNs retained for longer periods, and easing the computational burden of detailed subnational model calibration for long-term, spatially-tailored malaria control strategies in Burkina Faso and Nigeria.

The focus of Part I has been on malaria transmission modeling, harnessing geospatial products to inform and contextualize a mechanistic model. In Part II, this focus will switch, showcasing a strategy in which mechanistic modeling acts as a contextual backbone for geospatial techniques. Specifically, Part II will focus on a model framework for high-resolution estimates of ITN coverage in Africa over time. Because of the unique ways in which ITNs are procured, distributed, and utilized, appropriately capturing ITN coverage in both space and time required a mixed mechanistic and geospatial strategy. This framework provides another example of the many ways these two modeling methods can be creatively and generatively combined.

2.5 Additional Plots and Tables

The plots below show maps and covariate summaries for cluster counts of three to 14. In the time series, solid colored lines represent the median across the archetype, shaded areas indicate the interquartile range, and dotted lines indicate the 95% variance interval. Solid black lines represent the climate values of the representative site for each archetype, also indicated as black crosses on the map. For the six-archetype plot, dashed black lines and circles on the map represent the manually-selected sites that were used in the modeling analysis. Doughnut plots show the relative vector abundance of representative sites. Absent doughnuts indicate no mosquitoes in that site.

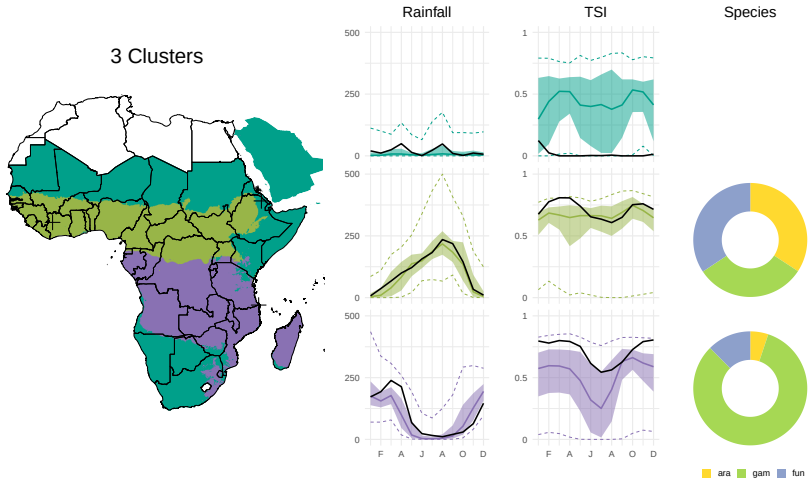


Figure 2.13: Maps and variable summaries for a cluster count of 3.

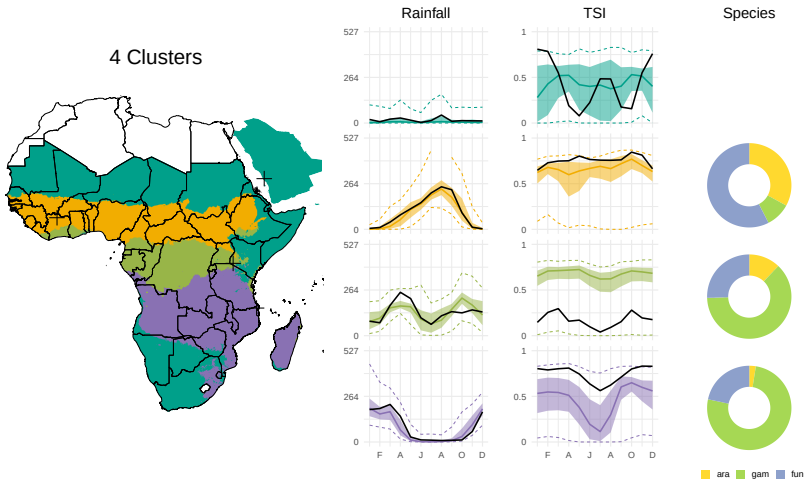


Figure 2.14: Maps and variable summaries for a cluster count of 4.

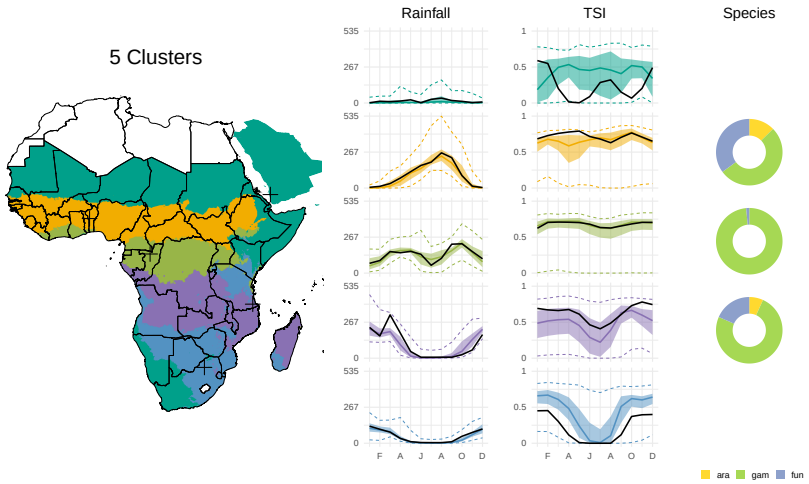


Figure 2.15: Maps and variable summaries for a cluster count of 5.

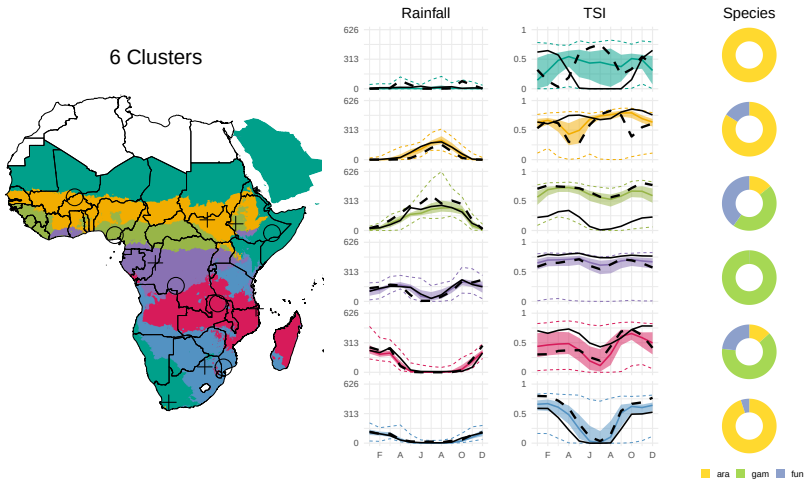


Figure 2.16: Maps and variable summaries for a cluster count of 6.

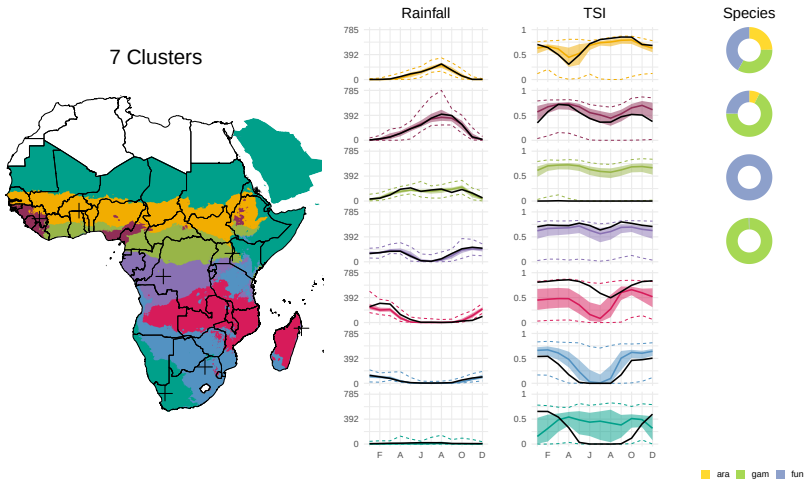


Figure 2.17: Maps and variable summaries for a cluster count of 7.

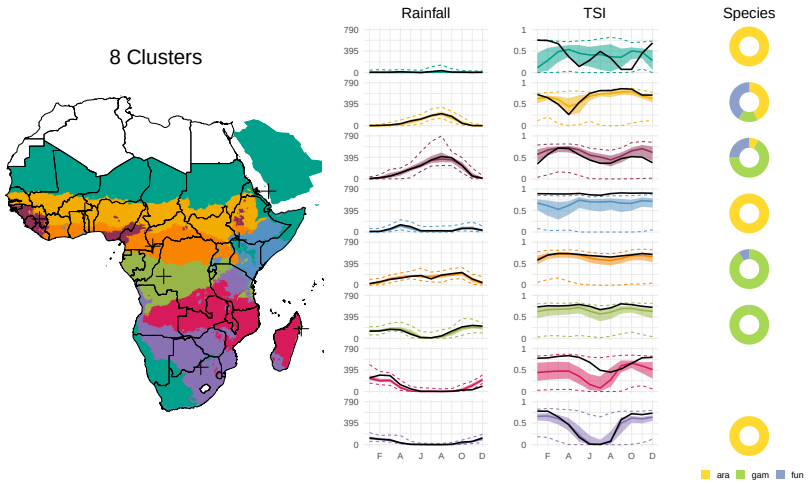


Figure 2.18: Maps and variable summaries for a cluster count of 8.

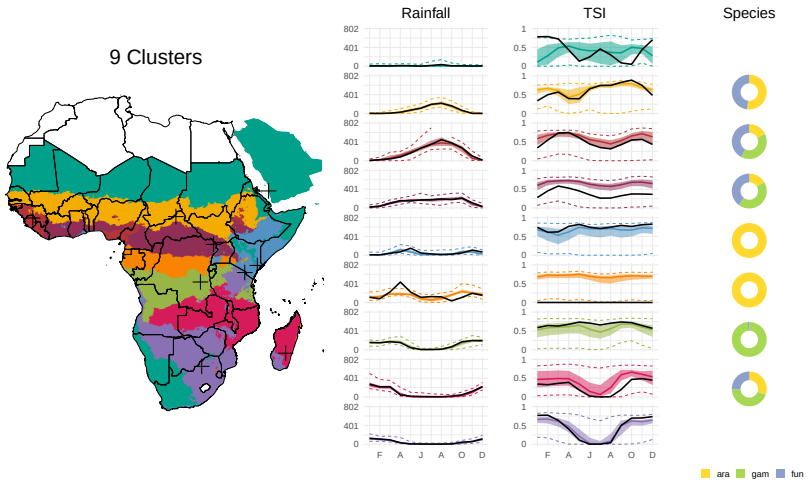


Figure 2.19: Maps and variable summaries for a cluster count of 9.

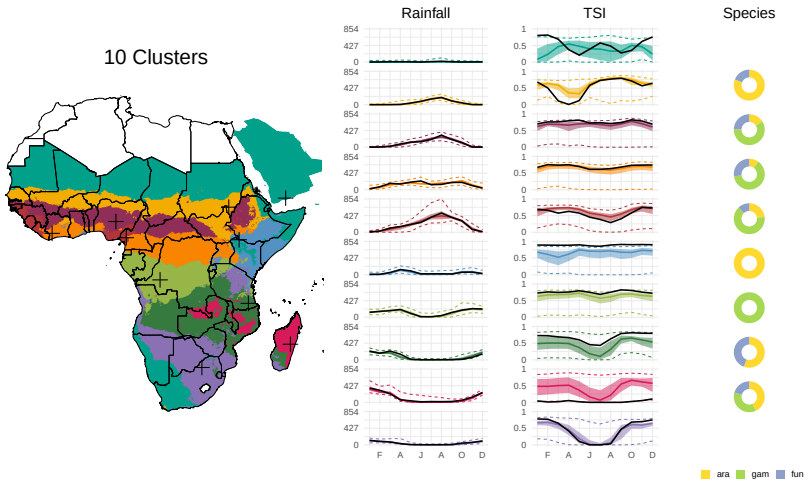


Figure 2.20: Maps and variable summaries for a cluster count of 10.

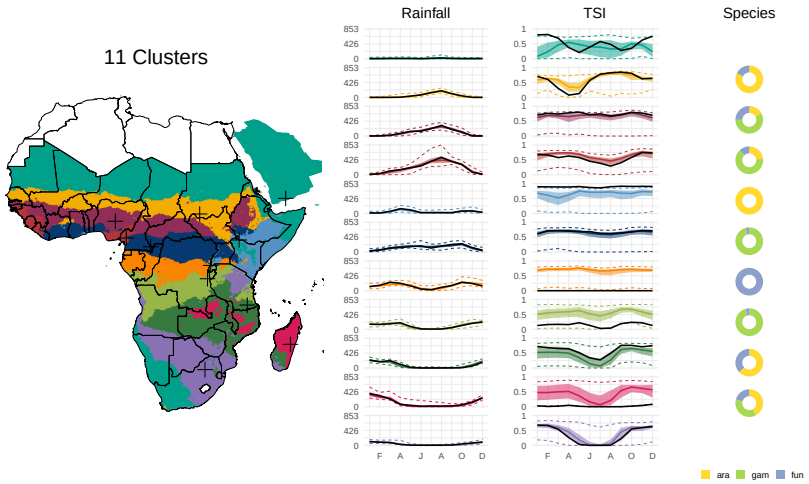


Figure 2.21: Maps and variable summaries for a cluster count of 11.

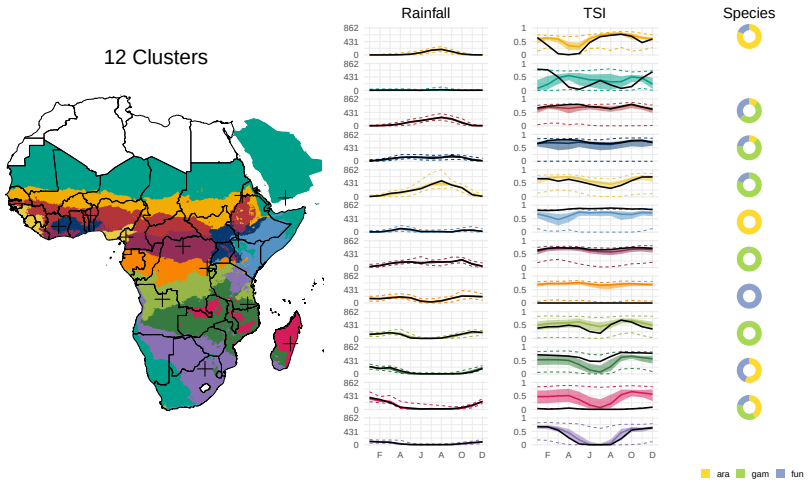


Figure 2.22: Maps and variable summaries for a cluster count of 12.

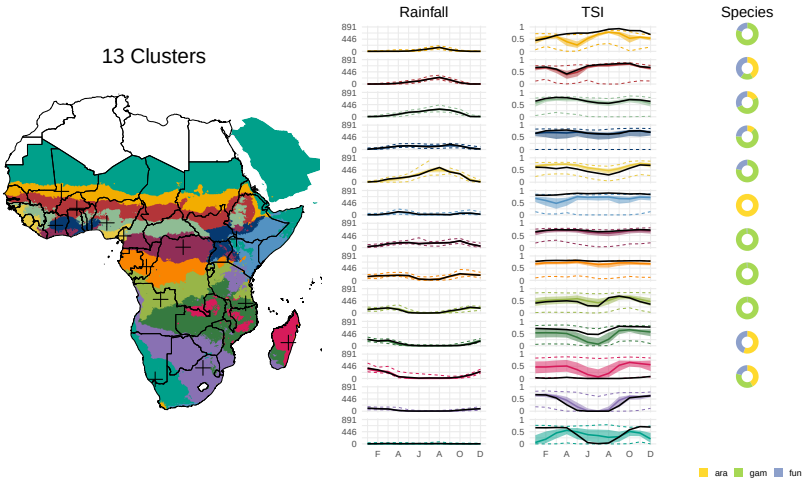


Figure 2.23: Maps and variable summaries for a cluster count of 13.

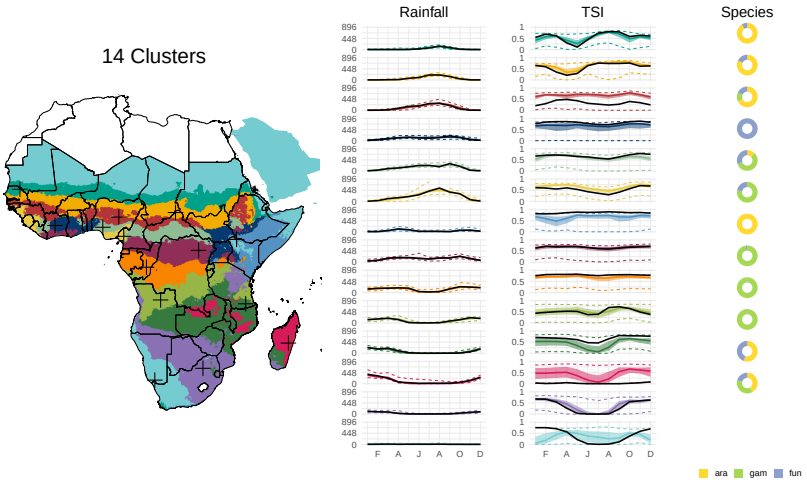


Figure 2.24: Maps and variable summaries for a cluster count of 14.

Table 2.4 shows the difference between the methods described in Chapter 1 and Chapter 2 in terms of population-at-risk (PAR). ITN, insecticide-treated net; AL CM, artemether-lumefantrine case management; IRS, indoor residual spraying; mAB, monoclonal antibodies; TBV, transmission-blocking vaccine; PEV, pre-erythrocytic vaccine; DP CM, dihydroartemisinin-piperaquine case management; ATSB, attractive targeted sugar bait.

Intervention descriptions and methods comparison.

Int #	**Description**	**Old PAR (Millions)**	**New PAR (Millions)**	**Difference (Millions)**	**Difference (%)**
0	Baseline	1359	1359	0	0
1	0/0/0% ITN/IRS/AL CM	1300	1319	-19	-1.5
2	20/0/0% ITN/IRS/AL CM	1161	1245	-84	-7.2
3	40/0/0% ITN/IRS/AL CM	999	1054	-55	-5.5
4	60/0/0% ITN/IRS/AL CM	783	846	-63	-8.1
5	80/0/0% ITN/IRS/AL CM	523	588	-65	-12.4
6	0/20/0% ITN/IRS/AL CM	1049	1105	-56	-5.3
7	20/20/0% ITN/IRS/AL CM	843	965	-122	-14.5
8	40/20/0% ITN/IRS/AL CM	661	722	-61	-9.2
9	60/20/0% ITN/IRS/AL CM	500	548	-48	-9.6
10	80/20/0% ITN/IRS/AL CM	352	370	-18	-5.1
11	0/40/0% ITN/IRS/AL CM	699	789	-90	-12.9
12	20/40/0% ITN/IRS/AL CM	532	617	-85	-16.0
13	40/40/0% ITN/IRS/AL CM	413	455	-42	-10.2
14	60/40/0% ITN/IRS/AL CM	305	335	-30	-9.9
15	80/40/0% ITN/IRS/AL CM	207	236	-29	-14.0
16	0/60/0% ITN/IRS/AL CM	429	484	-55	-12.8
17	20/60/0% ITN/IRS/AL CM	340	357	-17	-5.0
18	40/60/0% ITN/IRS/AL CM	279	274	5	1.8
19	60/60/0% ITN/IRS/AL CM	195	195	0	0
20	80/60/0% ITN/IRS/AL CM	156	158	-2	-1.3
21	0/80/0% ITN/IRS/AL CM	234	261	-27	-11.5
22	20/80/0% ITN/IRS/AL CM	196	209	-13	-6.6
23	40/80/0% ITN/IRS/AL CM	160	164	-4	-2.5
24	60/80/0% ITN/IRS/AL CM	130	125	5	3.9
25	80/80/0% ITN/IRS/AL CM	109	106	3	2.8
26	0/0/20% ITN/IRS/AL CM	1203	1208	-5	-0.4
27	20/0/20% ITN/IRS/AL CM	1022	1056	-34	-3.3
28	40/0/20% ITN/IRS/AL CM	808	881	-73	-9.0
29	60/0/20% ITN/IRS/AL CM	580	647	-67	-11.6

Continued on next page

Intervention descriptions and methods comparison, cont.					
Int #	**Description**	**Old PAR (Millions)**	**New PAR (Millions)**	**Difference (Millions)**	**Difference (%)**
30	80/0/20% ITN/IRS/AL CM	423	449	-26	-6.1
31	0/20/20% ITN/IRS/AL CM	912	953	-41	-4.5
32	20/20/20% ITN/IRS/AL CM	733	758	-25	-3.4
33	40/20/20% ITN/IRS/AL CM	534	575	-41	-7.7
34	60/20/20% ITN/IRS/AL CM	390	417	-27	-6.9
35	80/20/20% ITN/IRS/AL CM	275	271	4	1.5
36	0/40/20% ITN/IRS/AL CM	569	628	-59	-10.4
37	20/40/20% ITN/IRS/AL CM	422	469	-47	-11.1
38	40/40/20% ITN/IRS/AL CM	333	331	2	0.6
39	60/40/20% ITN/IRS/AL CM	239	244	-5	-2.1
40	80/40/20% ITN/IRS/AL CM	183	183	0	0
41	0/60/20% ITN/IRS/AL CM	335	359	-24	-7.2
42	20/60/20% ITN/IRS/AL CM	257	259	-2	-0.8
43	40/60/20% ITN/IRS/AL CM	202	195	7	3.5
44	60/60/20% ITN/IRS/AL CM	155	147	8	5.2
45	80/60/20% ITN/IRS/AL CM	117	113	4	3.4
46	0/80/20% ITN/IRS/AL CM	195	194	1	0.5
47	20/80/20% ITN/IRS/AL CM	166	147	19	11.5
48	40/80/20% ITN/IRS/AL CM	128	116	12	9.4
49	60/80/20% ITN/IRS/AL CM	108	103	5	4.6
50	80/80/20% ITN/IRS/AL CM	93	82	11	11.8
51	0/0/40% ITN/IRS/AL CM	1026	1053	-27	-2.6
52	20/0/40% ITN/IRS/AL CM	828	870	-42	-5.1
53	40/0/40% ITN/IRS/AL CM	641	674	-33	-5.2
54	60/0/40% ITN/IRS/AL CM	475	490	-15	-3.2
55	80/0/40% ITN/IRS/AL CM	329	331	-2	-0.6
56	0/20/40% ITN/IRS/AL CM	712	760	-48	-6.7
57	20/20/40% ITN/IRS/AL CM	563	557	6	1.1
58	40/20/40% ITN/IRS/AL CM	421	403	18	4.3
59	60/20/40% ITN/IRS/AL CM	300	289	11	3.7
60	80/20/40% ITN/IRS/AL CM	214	204	10	4.7
61	0/40/40% ITN/IRS/AL CM	444	441	3	0.7
62	20/40/40% ITN/IRS/AL CM	327	321	6	1.8
63	40/40/40% ITN/IRS/AL CM	241	236	5	2.1
64	60/40/40% ITN/IRS/AL CM	182	168	14	7.7
65	80/40/40% ITN/IRS/AL CM	155	122	33	21.3
66	0/60/40% ITN/IRS/AL CM	247	252	-5	-2.0
67	20/60/40% ITN/IRS/AL CM	199	186	13	6.5
68	40/60/40% ITN/IRS/AL CM	163	138	25	15.3
69	60/60/40% ITN/IRS/AL CM	121	112	9	7.4
70	80/60/40% ITN/IRS/AL CM	100	90	10	10.0
71	0/80/40% ITN/IRS/AL CM	155	137	18	11.6

Continued on next page

Intervention descriptions and methods comparison, cont.					
Int #	**Description**	**Old PAR (Millions)**	**New PAR (Millions)**	**Difference (Millions)**	**Difference (%)**
72	20/80/40% ITN/IRS/AL CM	130	112	18	13.9
73	40/80/40% ITN/IRS/AL CM	110	93	17	15.5
74	60/80/40% ITN/IRS/AL CM	95	74	21	22.1
75	80/80/40% ITN/IRS/AL CM	84	63	21	25.0
76	0/0/60% ITN/IRS/AL CM	824	835	-11	-1.3
77	20/0/60% ITN/IRS/AL CM	670	662	8	1.2
78	40/0/60% ITN/IRS/AL CM	497	477	20	4.0
79	60/0/60% ITN/IRS/AL CM	362	330	32	8.8
80	80/0/60% ITN/IRS/AL CM	247	215	32	13.0
81	0/20/60% ITN/IRS/AL CM	549	543	6	1.1
82	20/20/60% ITN/IRS/AL CM	419	397	22	5.3
83	40/20/60% ITN/IRS/AL CM	311	289	22	7.1
84	60/20/60% ITN/IRS/AL CM	216	196	20	9.3
85	80/20/60% ITN/IRS/AL CM	164	135	29	17.7
86	0/40/60% ITN/IRS/AL CM	330	310	20	6.1
87	20/40/60% ITN/IRS/AL CM	246	216	30	12.2
88	40/40/60% ITN/IRS/AL CM	184	163	21	11.4
89	60/40/60% ITN/IRS/AL CM	143	124	19	13.3
90	80/40/60% ITN/IRS/AL CM	116	89	27	23.3
91	0/60/60% ITN/IRS/AL CM	188	164	24	12.8
92	20/60/60% ITN/IRS/AL CM	147	130	17	11.6
93	40/60/60% ITN/IRS/AL CM	127	102	25	19.7
94	60/60/60% ITN/IRS/AL CM	104	80	24	23.1
95	80/60/60% ITN/IRS/AL CM	87	65	22	25.3
96	0/80/60% ITN/IRS/AL CM	127	104	23	18.1
97	20/80/60% ITN/IRS/AL CM	105	83	22	21.0
98	40/80/60% ITN/IRS/AL CM	91	70	21	23.1
99	60/80/60% ITN/IRS/AL CM	76	60	16	21.1
100	80/80/60% ITN/IRS/AL CM	66	53	13	19.7
101	0/0/80% ITN/IRS/AL CM	656	620	36	5.5
102	20/0/80% ITN/IRS/AL CM	510	451	59	11.6
103	40/0/80% ITN/IRS/AL CM	374	319	55	14.7
104	60/0/80% ITN/IRS/AL CM	260	213	47	18.1
105	80/0/80% ITN/IRS/AL CM	181	140	41	22.7
106	0/20/80% ITN/IRS/AL CM	406	352	54	13.3
107	20/20/80% ITN/IRS/AL CM	300	247	53	17.7
108	40/20/80% ITN/IRS/AL CM	221	176	45	20.4
109	60/20/80% ITN/IRS/AL CM	161	123	38	23.6
110	80/20/80% ITN/IRS/AL CM	119	90	29	24.4
111	0/40/80% ITN/IRS/AL CM	225	189	36	16.0
112	20/40/80% ITN/IRS/AL CM	182	142	40	21.9
113	40/40/80% ITN/IRS/AL CM	140	107	33	23.6

Continued on next page

Intervention descriptions and methods comparison, cont.					
Int #	**Description**	**Old PAR (Millions)**	**New PAR (Millions)**	**Difference (Millions)**	**Difference (%)**
114	60/40/80% ITN/IRS/AL CM	111	81	30	27.0
115	80/40/80% ITN/IRS/AL CM	95	65	30	31.6
116	0/60/80% ITN/IRS/AL CM	140	108	32	22.9
117	20/60/80% ITN/IRS/AL CM	122	89	33	27.1
118	40/60/80% ITN/IRS/AL CM	98	73	25	25.5
119	60/60/80% ITN/IRS/AL CM	80	60	20	25.0
120	80/60/80% ITN/IRS/AL CM	67	52	15	22.4
121	0/80/80% ITN/IRS/AL CM	96	68	28	29.2
122	20/80/80% ITN/IRS/AL CM	84	62	22	26.2
123	40/80/80% ITN/IRS/AL CM	71	54	17	23.9
124	60/80/80% ITN/IRS/AL CM	63	48	15	23.8
125	80/80/80% ITN/IRS/AL CM	54	41	13	24.1
126	80/80/80% ITN/IRS/AL CM + 40% mAb	54	38	16	29.6
127	80/80/80% ITN/IRS/AL CM + 40% TBV	42	35	7	16.7
128	80/80/80% ITN/IRS/AL CM + 40% PEV	55	41	14	25.5
129	80/80/80% ITN/IRS/AL CM + 80% mAb	49	37	12	24.5
130	80/80/80% ITN/IRS/AL CM + 80% TBV	33	32	1	3.0
131	80/80/80% ITN/IRS/AL CM + 80% PEV	57	41	16	28.1
132	80/80/80% ITN/IRS/DP CM	54	42	12	22.2
133	0/0/0% ITN/IRS/AL CM + 0.15% ATSB Kill Rate	1264	1305	-41	-3.2
134	0/0/0% ITN/IRS/AL CM + 3% ATSB Kill Rate	381	423	-42	-11.0
135	0/0/0% ITN/IRS/AL CM + 15% ATSB Kill Rate	12	13	-1	-8.3
136	0/0/0% ITN/IRS/AL CM + 25% ATSB Kill Rate	4	0	4	100.0
137	20/20/20% ITN/IRS/AL CM + 0.15% ATSB Kill Rate	637	712	-75	-11.8
138	20/20/20% ITN/IRS/AL CM + 3% ATSB Kill Rate	125	113	12	9.6
139	20/20/20% ITN/IRS/AL CM + 15% ATSB Kill Rate	10	9	1	10.0
140	20/20/20% ITN/IRS/AL CM + 25% ATSB Kill Rate	0	0	0	0

Continued on next page

Intervention descriptions and methods comparison, cont.					
Int #	**Description**	**Old PAR (Millions)**	**New PAR (Millions)**	**Difference (Millions)**	**Difference (%)**
141	40/40/40% ITN/IRS/AL CM + 0.15% ATSB Kill Rate	225	206	19	8.4
142	40/40/40% ITN/IRS/AL CM + 3% ATSB Kill Rate	56	49	7	12.5
143	40/40/40% ITN/IRS/AL CM + 15% ATSB Kill Rate	7	10	-3	-42.9
144	40/40/40% ITN/IRS/AL CM + 25% ATSB Kill Rate	0	0	0	0
145	60/60/60% ITN/IRS/AL CM + 0.15% ATSB Kill Rate	96	75	21	21.9
146	60/60/60% ITN/IRS/AL CM + 3% ATSB Kill Rate	31	28	3	9.7
147	60/60/60% ITN/IRS/AL CM + 15% ATSB Kill Rate	5	0	5	100.0
148	60/60/60% ITN/IRS/AL CM + 25% ATSB Kill Rate	0	0	0	0
149	80/80/80% ITN/IRS/AL CM + 0.15% ATSB Kill Rate	53	39	14	26.4
150	80/80/80% ITN/IRS/AL CM + 3% ATSB Kill Rate	23	20	3	13.0
151	80/80/80% ITN/IRS/AL CM + 15% ATSB Kill Rate	4	0	4	100.0
152	80/80/80% ITN/IRS/AL CM + 25% ATSB Kill Rate	0	0	0	0

Table 2.4: Interventions and methods comparisons. "Old" refers to the methods and results from Chapter 2, while "New" refers to methods and results from Chapter 1.

Bibliography

1. Nájera, J. A., González-Silva, M. & Alonso, P. L. Some Lessons for the Future from the Global Malaria Eradication Programme (1955–1969). *PLOS Medicine* **8,** e1000412. ISSN: 1549-1676. https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000412 (2011).
2. *Bill and Melinda Gates Call for New Commitment to Eradicate Malaria* Oct. 2007. https://philanthropynewsdigest.org/news/bill-and-melinda-gates-call-for-new-commitment-to-eradicate-malaria (2021).
3. Rabinovich, R. N. *et al.* malERA: An updated research agenda for malaria elimination and eradication. *PLOS Medicine* **14,** e1002456. ISSN: 1549-1676. https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002456 (2017).
4. WHO Strategic Advisory Group on Malaria Eradication. *Malaria eradication: benefits, future scenarios and feasibility.* tech. rep. (2019), 20. https://www.who.int/publications-detail/strategic-advisory-group-malaria-eradication-executive-summary.
5. Feachem, R. G. A. *et al.* Malaria eradication within a generation: ambitious, achievable, and necessary. *The Lancet* **394,** 1056–1112. ISSN: 0140-6736. http://www.thelancet.com/article/S0140673619311390/fulltext (2019).
6. Ton, J. F., Flaxman, S., Sejdinovic, D. & Bhatt, S. Spatial mapping with Gaussian processes and nonstationary Fourier features. *Spatial Statistics* **28,** 59–78. ISSN: 2211-6753 (2018).
7. Milton, P., Coupland, H., Giorgi, E. & Bhatt, S. Spatial analysis made easy with linear regression and kernels. *Epidemics* **29.** ISSN: 1878-0067. https://pubmed.ncbi.nlm.nih.gov/31561884/ (2019).
8. IPCC *et al. Climate Change 2014: Synthesis Report.* tech. rep. (2014). http://www.ipcc.ch..
9. CHIRPS. *CHIRPS: Rainfall Estimates from Rain Gauge and Satellite Observations* https://www.chc.ucsb.edu/data/chirps (2021).
10. Weiss, D. J. *et al.* Air temperature suitability for Plasmodium falciparum malaria transmission in Africa 2000-2012: A high-resolution spatiotemporal prediction. *Malaria Journal* **13.** ISSN: 14752875 (2014).

11. Sinka, M. E. *et al.* Modelling the relative abundance of the primary African vectors of malaria before and after the implementation of indoor, insecticide-based vector control. *Malaria Journal* **15,** 142. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/s12936-016-1187-8 (2016).

12. Chabot-Couture, G., Nigmatulina, K. & Eckhoff, P. An environmental data set for vector-borne disease modeling and epidemiology. *PLoS ONE* **9.** ISSN: 19326203 (2014).

13. Cameron, E. *et al.* Defining the relationship between infection prevalence and clinical incidence of Plasmodium falciparum malaria. *Nature Communications* **6,** 8170. ISSN: 2041-1723. http://www.nature.com/doifinder/10.1038/ncomms9170 (2015).

14. Nissan, H., Ukawuba, I. & Thomson, M. Climate-proofing a malaria eradication strategy. *Malar J* **20,** 190. https://doi.org/10.1186/s12936-021-03718-x (2021).

Part II

Estimating Insecticide-Treated Net Coverage

The following chapters present a second use case for the utility of combining mechanistic and geospatial models in malaria research, this time in the context of intervention coverage mapping. Like Part I, these chapters highlight how the flexibility of mechanistic modeling and the statistical strength of geospatial modeling complement each other to generate high-resolution maps of a complex phenomenon. However, while Part I focused primarily on methods development and implications, Part II focuses extensively on model results, how these results line up with other studies in the field, and the policy implications of these findings.

Chapter 3

Maps and Metrics of Insecticide-Treated Net Coverage in Africa: Access, Use, and Nets-Per-Capita, 2000-2020

> It says that when you build a thing you cannot merely build that thing in isolation, but must also repair the world around it, and within it, so that the larger world at that one place becomes more coherent, and more whole; and the thing which you make takes its place in the web of nature, as you make it.
>
> Christopher Alexander, Sara Ishikawa, & Murray Silverstein
> *A Pattern Language*

3.1 Introduction

Insecticide-treated nets (ITNs) are one of the most widespread, impactful, and cost-effective tools for combating malaria transmission in Africa, averting an estimated 450 million cases from 2000 to 2015 [1–3]. Mass ITN distributions have been a core

component of malaria control for almost two decades, over which period billions of dollars of aid have flowed into ITN production and distribution, resulting in the delivery of more than two billion nets to households across the continent. While the broad importance of ITNs is generally uncontested, many questions remain regarding optimal net allocation strategies, net owners' decisions on when to use and discard nets, and how to position ITN campaigns within a complex landscape of risk, cost, and impact. While others have addressed these questions locally and specifically, this analysis aggregates data and fills gaps across space and time to provide context and detect patterns in ITN distribution, utilization, and retention for 40 malaria–endemic countries in Africa from 2000 to 2020. Crucial to this modeling strategy is the combination of a mechanistic commodity supply chain model with a geospatial regression model.

In summarising the coverage of ITN distributions, two indicators are generally used: access, which measures what proportion of the population could sleep under a net assuming two people per net [4]; and use, which measures what proportion of the population does sleep under a net. These two metrics are often combined to calculate the use rate, defined as use among those with access. A final indicator, Nets-Per-Capita (NPC), tracks net volume among the population and is primarily used for procurement purposes [5, 6]. Throughout this chapter, metrics are either referred to specifically by name, or the term "coverage" is used to indicate any combination of metrics. All results in this analysis are reported among the population at risk of malaria, defined here as anyone living in a zone with stable *Plasmodium falciparum* malaria transmission [7].

Initially, ITN interventions were targeted at children under five and pregnant women [8]. However, targets for ITN coverage have since expanded to include the entire population at risk [9], though "universal coverage" targets are commonly set to 80% for both access and use [5, 10]. In recent years, language around coverage

targets has softened to acknowledge the heterogeneous intervention needs of different settings, though universal coverage remains the official policy [11].

Together, these coverage metrics encapsulate three key challenges to optimal ITN performance: distribution, utilization, and retention. Distributing nets from manufacturers to homes across Africa presents a massive supply-chain and commodity challenge that hinders sufficient net acquisition for many households. At higher coverage levels, there is some evidence of ITNs being overallocated to more accessible households that already have sufficient nets, leaving coverage gaps for less protected households [4, 12]. While ITN utilization is often high among those with access [13–16], heterogeneities have been observed across space [14], seasons [16], age groups [17, 18], and genders [19]. There is some evidence of a small proportion of nets being misused for fishing [20–22] and older nets being repurposed [23], but these activities do not occur at a sufficient scale to cause concern from a malaria control perspective. Finally, many studies have suggested that median net retention times may be dramatically shorter than the three-year duration presumed by mass campaign schedules [24–27], though others have estimated longer survival times [28–30]. Understanding the history and spatial distribution of ITN access, use, and NPC is crucial for effective ITN policy planning.

This analysis utilizes a Bayesian mixed modeling framework built upon data from net manufacturers, national programs, and cross-sectional household surveys over the past 20 years to estimate the history of ITN coverage metrics in 40 sub-Saharan African countries. This approach includes two main steps (Figure 3.1). First, a national-level "stock and flow" mechanistic model tracks the distribution, acquisition, and loss of ITNs by triangulating data from the three sources listed above. This step estimates both ITN retention times and ITN crop, the total number of nets in the community. Second, a series of geospatial models disaggregates this national time trend down to 5km-by-5km pixel resolution by estimating ITN access

deviation, the local variation from national mean access, and use gap, the local difference between access and use. From access and use, the ITN use rate, or use among those with access, can be determined. NPC is calculated similar to access via a nets-per-capita deviation model.

Previous publications have reported on ITN coverage metrics over time [12, 31] or space [1, 14, 32, 33]. This analysis presents high-resolution maps of ITN access, use, and NPC from 2000 to 2020. Findings support the evidence that ITN use is high among those with access, but that insufficient access continues to hinder progress towards coverage targets. Results show that universal ITN access is difficult to achieve due to the volume of nets needed and is subsequently difficult to maintain due to shorter-than-expected median net retention times. As an additional challenge, the relationship between NPC and access shows saturation effects which make high levels of access extremely costly to achieve. Resolving these allocative inefficiencies could dramatically increase ITN access in countries with high per-capita net coverage. In addition to reporting access, use, and use rate independently, this chapter includes the results of a relative gain analysis to quantify the improvements in ITN use that could be achieved by improving access versus improving the use rate. These maps and national estimates may be useful for benchmarking and comparison exercises, and are publicly available for use as covariates in other malaria-related analyses.

3.2 Definitions

Net Types

Pre-treated nets with an effective insecticidal lifespan of at least three years are defined as Long-Lasting Insecticide-Treated Nets, or "LLINs". Treated nets obtained or re-treated within the past 12 months are defined as Conventional Insecticide-

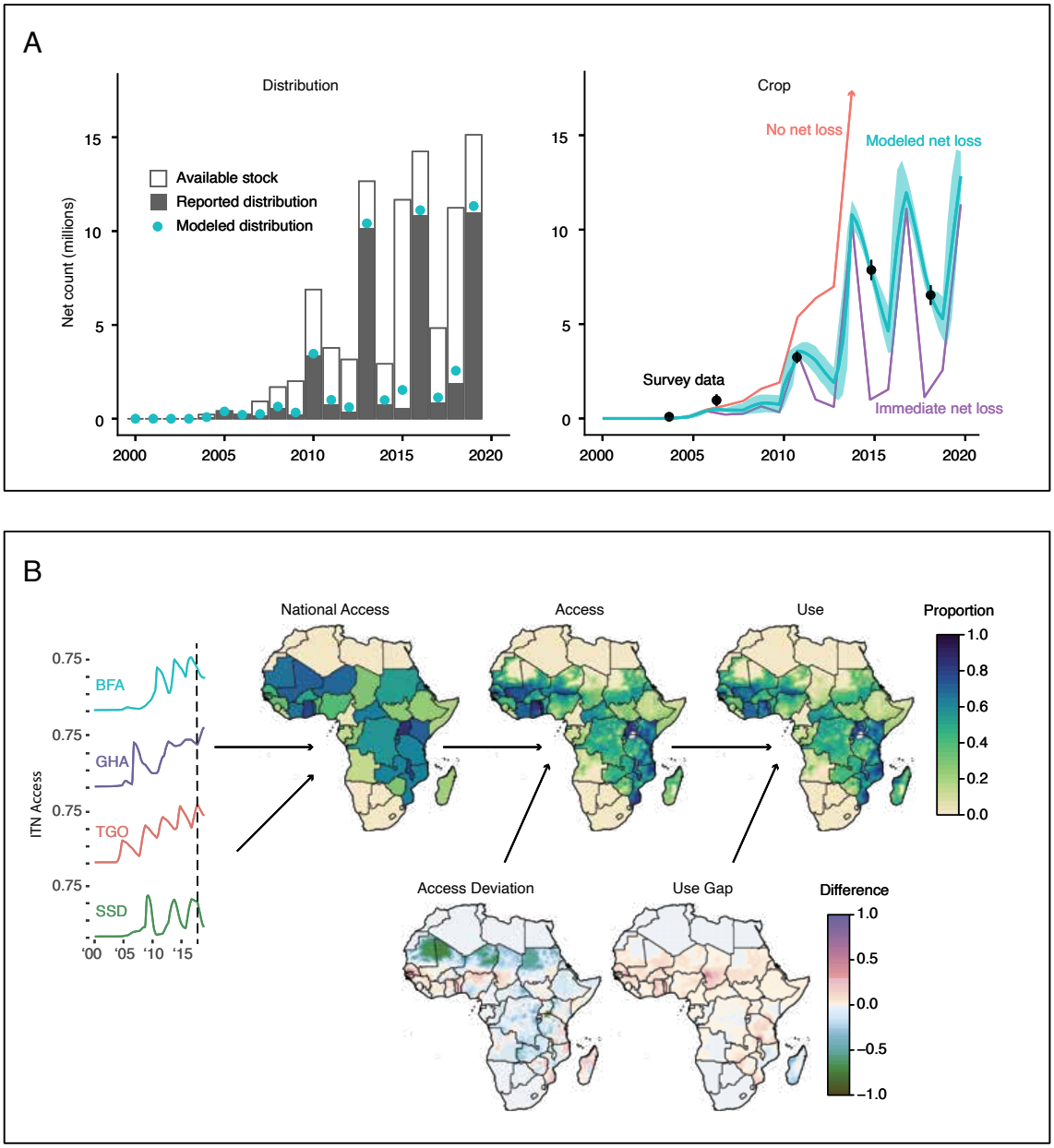


Figure 3.1: Insecticide-Treated Net (ITN) model summaries. A. Mechanistic "stock and flow" model. For each country (Burkina Faso shown for reference), the number of nets distributed must be no less than the reported distribution count and no more than the available stock (solid bars, left). Net loss follows an S-shaped curve whose steepness is fitted according to survey data (right). If nets were never discarded, net crop would increase cumulatively with every distribution (red line), whereas if nets were discarded immediately net crop would equal net distribution (purple line). The fitted curve (blue line and 95% credible interval) balances these two extremes. B. Geostatistical regression model. After net crop time series are converted to net access time series, geospatial regression models are run on the difference metrics of "access deviation" and "use gap". Final maps of ITN access are calculated by adding national access and access deviation, while final maps of ITN use are calculated by adding access and use gap. Maps of nets-per-capita are calculated similarly to access. BFA: Burkina Faso; GHA: Ghana; TGO: Togo; SSD: South Sudan. These four access time series are shown as examples, but 40 countries are included in the analysis.

Treated Nets, or "cITNs". LLINs and cITNs collectively comprise "ITNs", or simply "nets", as this analysis does not consider untreated nets.

Net Movement

ITN "delivery" refers to manufacturer shipment of nets to national programs or other distributing bodies, while ITN "distribution" refers to the provision of nets to end users. Most countries have continuous ITN distribution channels through antenatal clinics and child immunization programs, which are supplemented every 3-4 years with mass distribution campaigns directly to households. Some countries distribute ITNs through schools or community-level systems in addition to or in place of mass campaigns. ITN "stock" refers to the number of nets available to distribute at a given time. Because programs may not immediately distribute all of the nets delivered to them, stock in a given year is not necessarily equal to reported manufacturer delivery counts.

ITN "crop" refers to the total number of nets in homes at a given time point. Crop depends upon both ITN distribution and "retention", the length of time for which nets are owned before being discarded or repurposed. Most countries plan mass net distributions every three years, under the assumption that average retention times are not much shorter than this [6, 34]. NPC is ITN crop divided by population-at-risk.

Net Coverage

A person is defined to have "access" to an ITN if they live in a household where they could sleep under an ITN, assuming two people per net [5]. Population-level access is the proportion of people with access. To avoid underestimating this metric, access is calculated at the individual level– i.e., in a household of ten people and three nets,

six people will be defined as "having access" even though the household as a unit does not have sufficient nets for all its inhabitants. This is in contrast to "household-level access", which captures the proportion of households that own at least one ITN for every two people [5]. Access and NPC are calculated from household net counts and household sizes in nationally-representative surveys. ITN "use" is the proportion of people who sleep under a net, measured in surveys by listing which people in a household slept under each net the night prior to the survey. For all survey-based metrics, the population denominator is calculated from "de facto household size", defined as the number of people who slept in the household the night prior to the survey. The "use rate", calculated as use divided by access, is the proportion of people with access to a net who slept under it. When calculated at the population level, all of these metrics use population-at-risk as a denominator, defined as anyone living in a zone with stable *Plasmodium falciparum* malaria transmission [7].

"Access deviation" is the difference between access in a specific location within a country and the national mean access. Access deviation is positive when local access is greater than the national mean, and negative when local access is below the national mean. The "use gap" is the difference between access and use in a given location. The use gap is positive when not everyone with access to a net sleeps under it, negative when more than two people sleep under a single net, and zero when everyone who has access to a net sleeps under it respecting the "two people per net" guideline. Similar to access deviation, NPC deviation is the difference between NPC in a specific location within a country and the national mean NPC.

Household Size

For surveys with data available at the household level, "household size" refers to the number of people who slept in the household the night prior to the survey. For the 33 surveys (Table 3.13) for which only aggregated measures were available, "mean

household size" references the survey-specific definition of the term.

3.3 Methods: Stock and Flow

The goal of this analysis is the creation of high-resolution maps of ITN access, use, and NPC. Typically, such maps are generated via geospatial regression techniques alone, whereas this analysis first uses a mechanistic model to calculate mean national-level trends of access and NPC before regressing on access deviation, use gap, and NPC deviation (Figure 3.1). The extra step of mechanistic modeling is necessary for two reasons. First, to incorporate non-survey data on net delivery and distribution that are crucial to constructing an accurate time series but would be challenging to incorporate into the traditional geospatial framework where time is handled via covariance structures. Second, because the primary outcome variables are often discontinuous in space or time, while regression techniques assume a smoothly-varying output metric. By detrending the data, this analysis generates outcome variables with the appropriate mathematical properties.

3.3.1 Background

The goal of the stock and flow model is to estimate ITN crop and access over time at the national level from 2000 to 2020 for 40 sub-Saharan African countries. The model triangulates and ensures consistency between three data sources: reports from net manufacturers about how many ITNs were delivered to national programs, reports from national programs about how many ITNs were distributed to households, and data extracted from household surveys about how many ITNs were found in houses. The first two data sources are annual; the third is cross-sectional and was observed between one and ten times in each country across the time period of interest.

The difficulty of estimating ITN crop from surveys alone, and the utility of a mechanistic model for incorporating added information from delivery and distribution data, was first introduced by Flaxman et al. in 2010 [31]. This paper established the fundamental framework for a "stock and flow" model of tracking net crop in a given country over time, but was constrained by several limitations. First, net crop was estimated with an annual time step, diluting the 3-4 month time span captured by any nationally representative survey and yielding relatively coarse estimates of net crop over time. Second, rates of net loss were estimated separately in the first, second, and third years of net life, with all nets presumed discarded by the end of three years – a parameterization which makes description of net retention difficult and which may ignore long-term net ownership. Third, the model reported "coverage" in units of "household ownership of at least one ITN" and "ITN use in children under 5", neither of which are currently recommended as evaluation metrics (though both are still often reported and used for policy decisions).

In 2015, Bhatt et al. [12] published an improved version of the stock and flow model featuring a quarterly time step, a smooth-compact loss function for ITN retention, and the explicit calculation of five different ITN metrics, including household ownership, household access, population access, population use, and the "ownership gap". Limitations of this model included unrealistically high coverage estimates in countries with no available survey data (most notably South Sudan). Additionally, while the decisions to use splines for missing net distribution data imputation and fit two time-varying parameters for the net loss function added an immense amount of flexibility to the model, it concurrently called into question the identifiability of the parameters being fit.

Neither the Flaxman nor Bhatt models reported on several quality assurance metrics that are now standard. First, although the number of surveys available for model fitting differs drastically from country to country (from zero to five in

Flaxman's models and zero to nine in Bhatt's), the sensitivity of model fit to survey inclusion is not reported with either model. Second, while both models published extensive documentation, neither group published their code or otherwise made the details of their analysis available for external review or replication.

The present analysis substantially updates and refines the Bhatt model, including:

- Simplified imputation and loss function parameterization to improve identifiability;
- Sensitivity analysis on survey inclusion (Section 3.3.4);
- Updated survey data across the full time series, including previously overlooked surveys from Ethiopia, South Sudan, and Zimbabwe;
- A fully refactored, publicly available codebase with Google Cloud functionality (https://github.com/bertozzivill/map-itn-cube/tree/publication-2021).

The only other mechanistic model of national-level ITN crop known to the author is NetCALC (https://www.vector-works.org/resources/netcalc-planning-tool/), a publicly-available spreadsheet-based model designed to assist countries with net procurement calculations. NetCALC is widely used by national programs and net durability studies [24, 25, 35, 36], but is not designed for the type of historical analysis or spatial disaggregation required here.

3.3.2 Data

Net Stock: LLIN Manufacturer Reports

Net stock data were obtained from the Alliance for Malaria Prevention's (AMP) Net Mapping Project (https://netmappingproject.allianceformalariaprevention.com/) via the World Health Organization (WHO) (private correspondence), and represent the number of LLINs delivered by manufacturers to countries annually from 2000 to 2019. This source includes LLINs donated by Global Fund, the President's Malaria Initiative (PMI), UNICEF, the Against Malaria Foundation, World Bank, UNITAID, the UK Department for International Development, the Canadian International Development Agency, private sales, and other sources. These data come directly to AMP from all WHO-approved net manufacturers, and were thus assumed to be highly complete. Data were not available for LLINs in 2020 and do not include cITNs at all, so for these cases the number of nets delivered was assumed to equal the number of nets distributed (see below).

Net Distribution: National Malaria Control Programs, the African Leaders Malaria Alliance, and the President's Malaria Initiative Malaria Operational Plans

The following data sources were available for ITN distributions at the national level:

- National Malaria Control Program (NMCP), via WHO: Distribution numbers, disaggregated by net type, from 2000 to 2018, with extensive missing values.
- African Leaders Malaria Alliance (ALMA, https://alma2030.org/): Mass net distribution numbers from 2016 to 2020. These data contain missing values in country-years without mass campaigns. The 2020 distribution numbers were

obtained in late September of 2020, and enumerate both nets already delivered and those projected to be delivered before the end of the year. Both of these channels were included in the model as "delivered".

- PMI Malaria Operational Plans (PMI MOPs, https://www.pmi.gov/resource-library/mops/fy-2020): Twenty-four countries in this analysis receive funding from PMI, and therefore publish annual malaria operational plans which include a three-year assessment of retrospective and prospective ITN need and distribution. For example, the 2019 Tanzanian MOP contains ITN distribution counts for 2018, and proposed distributions for 2019, and 2020. Some reports are later supplemented with a revised funding table giving updated distribution estimates. While these reports are generated by PMI, they include estimates of net distributions from other large in-country donors. These are a less accurate resource than NMCP or ALMA data as they are largely prospective, but provide useful information in the absence of other data.

To create a single cohesive time series of net distributions from 2000 to 2019, data from these sources were combined as follows (Figure 3.2):

1. From 2000 to 2015, NMCP data were used, with the process of missing value imputation described below.

2. From 2016 to 2020, where available, the higher of the NMCP or ALMA values were used.

3. From 2016 to 2020, where neither NMCP nor ALMA data were available but a PMI MOP was present, the latest available PMI estimate was used.

4. If no data from any source was available in a country-year, the minimum NMCP distribution value from 2014-2018 was used. This value was selected as a proxy for a typical number of routinely-distributed nets in a country-year.

NMCP data prior to 2016 were incomplete and required imputation. Missing values were primarily in country-years where the expected number of nets was zero (i.e. for cITNs after 2013 or LLINs prior to 2007), but some countries had missing values at points in the time series when nonzero distributions would be expected.

For years prior to each country's mass adoption of LLINs or all years for cITNs, missing distribution counts were set to zero. For missing values in other years, Table 3.1 lists the case-specific strategies used to estimate a typical number of routinely distributed nets, avoiding mass distributions.

Country	**Year**	**Interpolation**
Chad	2012	Take the minimum of 2010-2013.
Cote d'Ivoire	2007-8	Interpolate between 2006 and 2009.
Cote d'Ivoire	2012	Take the minimum of 2010-2013.
DRC	2005	Interpolate between 2004 and 2006.
Mauritania	2007	Interpolate between 2006 and 2008.
Tanzania	2013-2016	Take values from the most recent available PMI MOPs.
Togo	2009	Take the minimum of 2008-2011.

Table 3.1: Custom, country-specific strategies for interpolating missing net distribution data. DRC: Democratic Republic of the Congo; PMI MOP: President's Malaria Initiative Malaria Operational Plan.

Net distribution counts summarize complex campaigns requiring the coordinated effort of thousands of health workers. As reflected in the variety and missingness of the available data, determining and reporting these values presents numerous logistical challenges. The values presented here may not reflect nets lost in the process of distribution, or nets acquired through the private market. To reflect this uncertainty, wide prior distributions were placed on ITN distribution counts in the model.

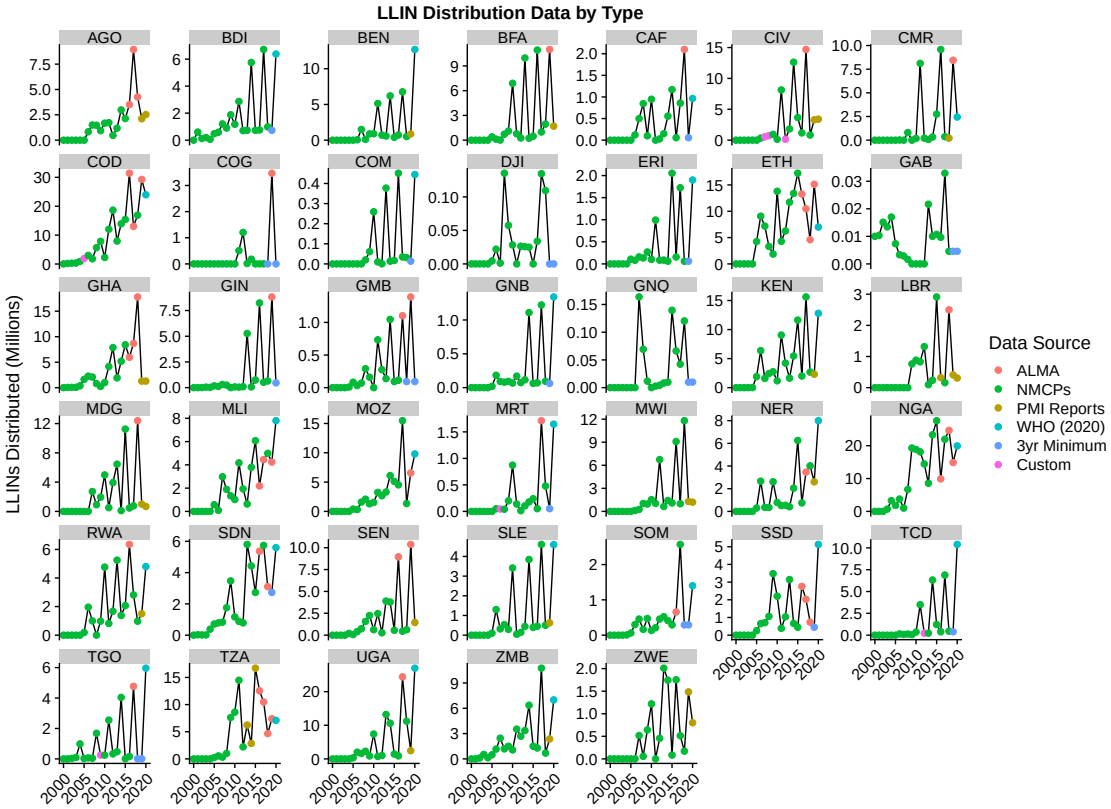


Figure 3.2: Long-lasting insecticide-treated net distribution by type, including direct data sources (red, green, gold, and cyan) and interpolated values (blue and pink). See Table 3.1 for more detail on the "Custom" strategy.

Net Crop: Nationally-Representative Surveys

To obtain data on net crop, ITN ownership and household size indicators were collated and extracted from 161 nationally-representative household surveys conducted in Sub-Saharan Africa from 2000 to 2019. These included the Demographic Health Survey (DHS), Malaria Indicator Survey (MIS), Multiple Indicator Cluster Survey (MICS) rounds 3, 4, and 5, AIDS Indicator Survey (AIS), and one Anemia and Parasitemia Survey (EA&P).

Data on the number of nets owned at the household level were available for 128 surveys in 36 countries (Table 3.2, Figure 3.3). Of these, most reported net type for every net owned in a household, while some reported net type only on one net per household. For those surveys that reported net type for all nets in the household, the number of ITNs owned was determined by summing ITNs by type. For those surveys where data on net type was only available for one net per household, the overall survey-level proportion of nets by type (non-ITN, cITN, or LLIN) was determined and multiplied by the number of nets in the surveyed household to estimate the number of LLINs and cITNs owned by each household. These household-level values were aggregated using the appropriate survey weights to generate national-level estimates of mean cITNs per household, mean LLINs per household, and mean household size. The standard errors for each of these metrics was used to inform the width of survey-specific prior distributions. Each survey was assigned to a single point in time, defined as the survey-weighted mean of the dates of household interviews.

For an additional 33 surveys in 20 countries, data were only available in aggregate from survey reports (Table 3.13). For these, measures of mean household size and mean cITNs and LLINs per household were collected from the relevant report tables. The standard errors reported for each of these metrics was used to inform the width

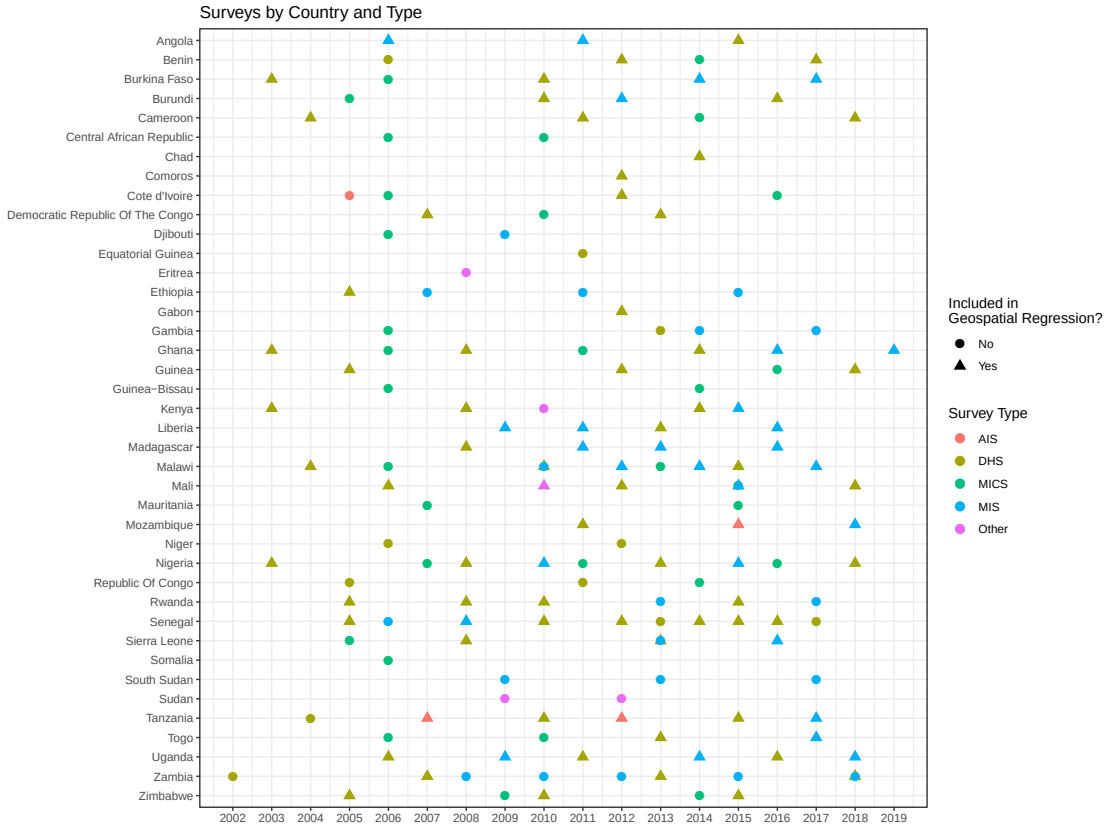


Figure 3.3: Nationally-representative surveys included in this analysis, listed by country, year, and survey type.

of survey-specific prior distributions. Each survey was assigned to a single point in time, defined as the midpoint between the beginning and end of data collection for the survey.

Population

Population and Population at Risk (PAR) are from the central MAP database. Population estimates combine information from WorldPop, AfriPop, and Institute for Health Metrics and Evaluation (IHME) population databases. For more details see the supplementary materials of Weiss et al. [37]. PAR was calculated by summing population in any pixel within the MAP-estimated *Plasmodium falciparum* transmission limits (downloadable at https://malariaatlas.org/explorer/#/).

Surveys used in ITN models.					
Country	**Survey Years**	**Source**	**No. of Clusters**	**No. of Individuals**	**Geospatial Data?**
Angola	2006-2007	MIS	115	13,952	Yes
Angola	2010-2011	MIS	238	39,951	Yes
Angola	2015-2016	DHS	625	72,870	Yes
Benin	2006	DHS	–	87,396	No
Benin	2011-2012	DHS	750	85,898	Yes
Benin	2014	MICS5	0	86,676	No
Benin	2017-2018	DHS	555	73,336	Yes
Burkina Faso	2003	DHS	400	58,844	Yes
Burkina Faso	2006	MICS3	–	37,070	No
Burkina Faso	2010	DHS	573	80,532	Yes
Burkina Faso	2014	MIS	252	38,393	Yes
Burkina Faso	2017-2018	MIS	245	36,307	Yes
Burundi	2005	MICS3	–	40,633	No
Burundi	2010-2011	DHS	376	40,983	Yes
Burundi	2012-2013	MIS	200	22,606	Yes
Burundi	2016-2017	DHS	554	76,528	Yes
Cameroon	2004	DHS	466	49,557	Yes
Cameroon	2011	DHS	578	70,294	Yes
Cameroon	2014	MICS5	–	51,031	No
Cameroon	2018-2019	DHS	430	58,474	Yes
Central African Republic	2006	MICS3	–	54,385	No
Central African Republic	2010	MICS4	–	52,355	No
Chad	2014-2015	DHS	624	96,005	Yes
Comoros	2012	DHS	252	23,580	Yes
Congo (Republic of)	2005	DHS	–	29,156	No
Congo (Republic of)	2011-2012	DHS	–	48,826	No
Congo (Republic of)	2014-2015	MICS5	–	67,159	No
Cote d'Ivoire	2005	AIS	–	23,475	No
Cote d'Ivoire	2006	MICS3	–	54,402	No
Cote d'Ivoire	2011-2012	DHS	351	49,278	Yes
Cote d'Ivoire	2016	MICS5	–	65,109	No
Democratic Republic of the Congo	2007	DHS	300	46,496	Yes
Democratic Republic of the Congo	2010	MICS4	–	58,510	No

Continued on next page

Surveys used in ITN models, cont.					
Country	**Survey Years**	**Source**	**No. of Clusters**	**No. of Individuals**	**Geospatial Data?**
Democratic Republic of the Congo	2013-2014	DHS	536	93,147	Yes
Djibouti	2006	MICS3	–	28,781	No
Djibouti	2009	MIS	–	22,373	No
Equatorial Guinea	2011	DHS	–	19,745	No
Eritrea	2008	MIS	–	8,814	No
Ethiopia	2005	DHS	535	64,914	Yes
Ethiopia	2007	MIS	–	32,380	No
Ethiopia	2011	MIS	–	47,248	No
Ethiopia	2015	MIS	–	53,335	No
Gabon	2012	DHS	336	40,597	Yes
Gambia	2005-2006	MICS3	–	44,877	No
Gambia	2013	DHS	–	50,347	No
Gambia	2014	MIS	–	42,633	No
Gambia	2017	MIS	–	40,393	No
Ghana	2003	DHS	412	25,498	Yes
Ghana	2006	MICS3	–	24,947	No
Ghana	2008	DHS	411	45,297	Yes
Ghana	2011	MICS4	–	52,969	No
Ghana	2014	DHS	427	42,292	Yes
Ghana	2016	MIS	200	22,332	Yes
Ghana	2019	MIS	200	23,000	Yes
Guinea	2005	DHS	295	36,950	Yes
Guinea	2012	DHS	300	43,876	Yes
Guinea	2016	MICS5	–	58,021	No
Guinea	2018	DHS	401	48,956	Yes
Guinea-Bissau	2006	MICS3	–	41,312	No
Guinea-Bissau	2014	MICS5	–	62,451	No
Kenya	2003	DHS	400	36,464	Yes
Kenya	2008-2009	DHS	398	37,790	Yes
Kenya	2010	MIS	–	26,946	No
Kenya	2014	DHS	1594	145,440	Yes
Kenya	2015	MIS	245	24,989	Yes
Liberia	2008-2009	MIS	150	21,876	Yes
Liberia	2011-2012	MIS	150	18,632	Yes
Liberia	2013	DHS	322	45,995	Yes
Liberia	2016	MIS	150	20,859	Yes
Madagascar	2008-2009	DHS	594	82,594	Yes
Madagascar	2011	MIS	267	39,337	Yes
Madagascar	2013	MIS	274	37,571	Yes
Madagascar	2016	MIS	358	47,640	Yes
Malawi	2004-2005	DHS	521	58,812	Yes

Continued on next page

Surveys used in ITN models, cont.					
Country	**Survey Years**	**Source**	**No. of Clusters**	**No. of Individuals**	**Geospatial Data?**
Malawi	2006	MICS3	–	131,021	No
Malawi	2010	DHS	849	115,027	Yes
Malawi		–	–	–	No
Malawi	2012	MIS	140	14,091	Yes
Malawi	2013-2014	MICS5	–	140,133	No
Malawi	2014	MIS	140	14,141	Yes
Malawi	2015-2016	DHS	850	117,833	Yes
Malawi	2017	MIS	150	16,330	Yes
Mali	2006	DHS	407	70,871	Yes
Mali	2010	EA & P	109	9,561	Yes
Mali	2012-2013	DHS	413	55,658	Yes
Mali	2015	MICS5	–	123,858	No
Mali	2015	MIS	177	38,705	Yes
Mali	2018	DHS	345	53,477	Yes
Mauritania	2007	MICS3	–	59,572	No
Mauritania	2015	MICS5	–	69,675	No
Mozambique	2011	DHS	610	61,159	Yes
Mozambique	2015	AIS	306	32,051	Yes
Mozambique	2018	MIS	224	28,126	Yes
Niger	2006	DHS	–	46,403	No
Niger	2012	DHS	–	61,680	No
Nigeria	2003	DHS	362	34,609	Yes
Nigeria	2007	MICS3	–	124,840	No
Nigeria	2008	DHS	886	154,946	Yes
Nigeria	2010	MIS	239	30,134	Yes
Nigeria	2011	MICS4	–	143,686	No
Nigeria	2013	DHS	896	176,949	Yes
Nigeria	2015	MIS	326	37,772	Yes
Nigeria	2016-2017	MICS5	–	189,319	No
Nigeria	2018	DHS	1389	186,327	Yes
Rwanda	2005	DHS	462	46,715	Yes
Rwanda	2007-2008	DHS	249	31,741	Yes
Rwanda	2010-2011	DHS	492	55,280	Yes
Rwanda	2013	MIS	–	20,450	No
Rwanda	2014-2015	DHS	492	53,770	Yes
Rwanda	2017	MIS	–	19,901	No
Senegal	2005	DHS	376	66,506	Yes
Senegal	2006	MIS	–	28,918	No
Senegal	2008-2009	MIS	320	104,488	Yes
Senegal	2010-2011	DHS	391	74,941	Yes
Senegal	2012-2013	DHS	200	39,756	Yes
Senegal	2013	DHS	–	37,518	No
Senegal	2014	DHS	200	39,447	Yes
Senegal	2015	DHS	214	41,042	Yes

Continued on next page

Surveys used in ITN models, cont.					
Country	**Survey Years**	**Source**	**No. of Clusters**	**No. of Individuals**	**Geospatial Data?**
Senegal	2016	DHS	214	40,573	Yes
Senegal	2017	DHS	–	77,084	No
Sierra Leone	2005	MICS3	–	42,719	No
Sierra Leone	2008	DHS	353	40,426	Yes
Sierra Leone	2013	DHS	435	74,290	Yes
Sierra Leone	2013	MIS	–	36,395	No
Sierra Leone	2016	MIS	336	39,580	Yes
Somalia	2006	MICS3	–	33,959	No
South Sudan	2009	MIS	–	17,001	No
South Sudan	2013	MIS	–	18,337	No
South Sudan	2017	MIS	–	32,358	No
Sudan	2009	MIS	–	30,423	No
Sudan	2012	MIS	–	26,547	No
Tanzania	2004-2005	DHS	–	46,516	No
Tanzania	2007-2008	AIS	475	85,034	Yes
Tanzania	2009-2010	DHS	475	47,357	Yes
Tanzania	2011-2012	AIS	583	50,996	Yes
Tanzania	2015-2016	DHS	608	61,030	Yes
Tanzania	2017	MIS	442	45,814	Yes
Togo	2006	MICS3	–	30,542	No
Togo	2010	MICS4	–	29,573	No
Togo	2013-2014	DHS	330	45,432	Yes
Togo	2017	MIS	171	22,478	Yes
Uganda	2006	DHS	368	43,396	Yes
Uganda	2009-2010	MIS	170	20,916	Yes
Uganda	2011	DHS	404	102,538	Yes
Uganda	2014-2015	MIS	210	26,464	Yes
Uganda	2016	DHS	696	88,174	Yes
Uganda	2018-2019	MIS	340	44,489	Yes
Zambia	2001-2002	DHS	–	35,941	No
Zambia	2007	DHS	319	33,794	Yes
Zambia	2008	MIS	–	20,430	No
Zambia	2010	MIS	–	20,042	No
Zambia	2012	MIS	–	16,928	No
Zambia	2013-2014	DHS	721	78,486	Yes
Zambia	2015	MIS	–	16,129	No
Zambia	2018	MIS	–	18,384	No
Zambia	2018-2019	DHS	545	62,342	Yes
Zimbabwe	2005-2006	DHS	398	41,289	Yes
Zimbabwe	2009	MICS3	–	52,194	No
Zimbabwe	2010-2011	DHS	406	40,711	Yes
Zimbabwe	2014	MICS5	–	67,536	No
Zimbabwe	2015-2016	DHS	400	41,894	Yes

Continued on next page

Surveys used in ITN models, cont.					
Country	**Survey Years**	**Source**	**No. of Clusters**	**No. of Individuals**	**Geospatial Data?**

Table 3.2: Surveys used in ITN models.

3.3.3 Model

Summary

National LLIN crop was estimated via a modified version of the mechanistic "stock and flow" model described below and in Bhatt et al. [12]. Briefly, a compartmental model was used to triangulate LLIN stock, distribution, and ownership data. Stock and distribution data were linked via a minimum function, mandating that no more than the available stock of nets could be distributed at a given time point. Distribution and ownership data were linked via a sigmoidal loss function to model discarding of LLINs over time. National cITN crop was modeled similarly, but no stock information was available for cITNs, so the upper bound on cITN distributions was omitted.

The model was run separately for each country. All parameters were fit separately by net type. After estimating LLIN and cITN crop, the two were summed into a single time series of national ITN crop. This was converted to ITN access via a household-size-based regression methodology described in Bhatt et al. [12] and below. All models were fitted in JAGS using the `rjags` package version 4.3.0, which produces draws from the posterior distributions of the fitted parameters. Each model generated 5,000 posterior draws, from which mean and 95% credible intervals were computed. To generate mean estimates and uncertainties for the conversion to access, 500 draws were randomly selected and propagated due to computational constraints.

For this analysis, the model described in Bhatt et al. [12] was refactored and rewritten in its entirety, but the model specification has remained largely constant. Key changes are:

- Updating input data from net manufacturers, NMCPs, and surveys, including all information from ALMA and PMI;
- Changing from a time-varying to a static loss function parameterization to improve model identifiability;
- Placing a cap on excess stock distribution to ensure a reliable fit to survey data.

A formal model specification is presented below. To clarify notation, the process is described for a single country and for LLINs only. The specification for cITNs is identical except where noted.

Stock and Flow

All symbols and their definitions are shown in the tables below. Values drawn directly from data are represented by Roman characters, while parameters derived through model fitting are represented by Greek characters. Sigmas (σ) refer to parameter-specific standard errors, and hats ($\hat{}$) over characters describe the modeled representations of data values.

Data on deliveries of LLINs to countries from manufacturers did not report uncertainty, but were assumed to be highly accurate. Therefore, LLIN deliveries m

Symbol	Meaning	Units	Notes
$Y, y \in Y$	Year count, year index	Years	
$Q, q \in Q$	Quarter count, quarter index	Quarters	$Q = Y * 4 + 1$; see text for details
t	Continuous time	Years	
$S, s \in S$	Survey count, survey index	Integer	
h	Household size	Individuals	
P	National population	Individuals	Source: MAP
m	Manufacturer deliveries of LLINs	Nets	Source: AMP
d	Reported cITN and LLIN distributions	Nets	Source: ALMA, NMCPs, and PMI
C, c^h, c^p	ITN crop (C), count per household c^h, and count per capita c^p	Nets	Source: Nationally representative surveys

Table 3.3: Stock and flow parameter definitions: indices and data.

in year y were modeled as:

$$\sigma_y^m \sim \text{Unif}(0, 0.075)$$

$$\hat{m}_y \sim N(m_y, m_y \sigma_y^m)$$

This relationship allows for a small amount of noise in the number of LLIN deliveries. Similar data were not available for cITNs.

Data on ITN distributions also did not report uncertainty, but were assumed to be somewhat less accurate than manufacturer data given the presence of missing values and the difficulties inherent in collecting such information (see Section 3.3.2). Therefore, modest prior distributions were placed on per-capita rates of ITN distribution:

$$\sigma_y^d \sim \text{Unif}(0, 0.03)$$

$$\hat{d}_y \sim N(\frac{d_y}{P_y}, \sigma_y^d) P_y$$

Symbol	**Meaning**	**Value**	**Units**
$\Gamma, \gamma^h, \gamma^p$	ITN crop (Γ), count per household γ^h, and count per capita γ^p		Nets
δ	Net distribution		Nets
α, ω	Initial and final LLIN stock for a given year		Nets
ϕ	Net distribution adjustment: proportion of surplus stock distributed	Beta(2,β) T(0,0.25)	Proportion
β	Shape parameter for Beta distribution of ϕ	Unif(20,24)	Unitless
κ	Rate parameter for net loss function	20	Unitless
τ	Time by which all nets are lost	Unif(5.5,20.7)	Years

Table 3.4: Stock and flow parameter definitions: modeled values $T(\cdot)$ indicates truncation.

Since no net delivery data was available for cITNs, $\hat{d}_y^{cITN}$ was used without adjustment, while $\hat{d}_y^{LLIN}$ was further adjusted using $\hat{m}_y$.

An LLIN time series was constructed using information about both stock and distributions. Every year, LLINs were delivered to a country (acquisition of stock) and distributed (dispersal of stock). Not all acquired stock was necessarily dispersed each year, allowing for the possibility of stock accumulation over time. Stock at the beginning and end of every year was tracked using parameters α_y and ω_y.

In a given year, no more LLINs could be distributed than the available stock α_y, and no fewer could be distributed than $\hat{d}_y$. If more than $\hat{d}_y$ LLINs were available, a small portion of additional stock could be distributed to allow a closer fit to survey data. However, the model should not distribute a large amount of additional stock, both in the interest of hewing to data on net distributions and because countries may receive shipments of nets in anticipation of distributions in the following year. Therefore, the proportion of surplus stock to distribute, ϕ_y, was modeled as a Beta distribution with shape parameters 2 and β_y, truncated at 0.25. A narrow prior distribution was placed on β such that the mean value of ϕ ranged between 0.07 and

0.09. The parameter ϕ can be thought of as a mechanistically-informed relaxation on the upper bound of the prior for LLIN distributions.

The LLIN distribution time series was constructed as follows. Beginning in year $y_0 = 2000$ and proceeding sequentially through each year:

$$\alpha_y = \hat{m}_{y_0} \quad \text{if} \quad y = y_0, \quad \text{otherwise} \quad \alpha_y = \omega_{y-1} + \hat{m}_y$$

$$\delta_y^{raw} = \min(\hat{m}_y, \hat{d}_y)$$

$$\beta_y \sim \text{Unif}(20, 24)$$

$$\phi_y \sim \text{Beta}(2, \beta_y) \quad \text{T}(0, 0.25)$$

$$\delta_y = \delta_y^{raw} + \phi_y(\alpha_y - \delta_y^{raw})$$

$$\omega_y = \alpha_y - \delta_y$$

where $T(\cdot)$ represents a truncated distribution. This yields an adjusted estimate for annual LLIN distributions, δ_y.

Next, annual ITN distributions δ_y were disaggregated the into quarterly net distributions δ_q. For each year y in Y:

1. Split the year into four quarters, $i \in 1:4$.
2. Define the proportion of ITNs distributed in each year-quarter, $p_{y,i} \sim \text{Unif}(0, 1)$, raked such that $\sum_i p_{y,i} = 1$.
3. Define $\delta_{y,i} = \delta_y p_{y,i}$ for each quarter i.
4. Reindex such that quarters are indexed as δ_q rather than $\delta_{y,i}$.

The calculation was performed separately by net type, yielding quarterly esti-

mates of net distribution δ_q^{LLIN} and δ_q^{cITN}. Since survey calibration requires interpolation between the beginning and end of each quarter, net distribution estimates were also calculated for the first quarter of year $Y + 1$. These were assumed to be the same as distributions in the final quarter of the last year for which distribution data were available, $\delta_Q = \delta_{Q-1}$. From these quarterly distribution counts, net loss can be tracked.

Following Bhatt et al. [12], a smooth-compact loss function for nets was used, with:

$$\text{Loss } (t, \kappa, \tau) = e^{\kappa - \kappa/(1-(t/\tau)^2)}$$

if $t < \tau$, and $\text{Loss}(t, \kappa, \tau) = 0$ otherwise.

κ is therefore a rate parameter, while τ dictates the time by which all nets will be lost. Previous versions of the model constrained κ and allowed τ to vary widely. To further aid identifiability, here κ was fixed at a value of 20 and τ was given a wide prior distribution $\tau \sim \text{Unif}(5.5, 20.7)$. This corresponds to net retention half-lives of between 1 and 3.79 years. These bounds were placed based on data from longitudinal studies, which rarely shows median retention times less than 1 or greater than 3.5 years [24–30, 36, 38, 39]. After model fitting, net retention half-life was calculated analytically as the time t at which $\text{Loss } (t, \kappa, \tau) = 0.5$. To estimate net loss, for each net type a $Q \times Q$ matrix A was constructed, where each column j represents a quarter of net distribution and row i tracks net retention in subsequent quarters. Entries of the matrix were populated according to the following rules:

- $A_{ij} = 0$ if $i < j$ (nets have yet to be distributed)
- $A_{ij} = \delta_j$ if $i = j$ (nets were distributed in that quarter)

- $A_{ij} = \delta_j \text{ Loss}(t, \kappa, \tau)$ if $i > j$, with $t = (i - j)\frac{1}{4}$.

Once A was populated, summing across columns generated estimated net crop by quarter, $\Gamma_q = \sum_{j \in 1:Q} A_{qj}$. Next, these quarterly distribution estimates were fit to the available survey data.

Each survey yielded an estimate of mean nets per household, c_s^h, and mean household size, h_s, each with standard errors. A national net crop C was estimated for each survey via $C_s = c_s^h \frac{P_y}{h_s}$, with P_y the national population in the year y in which survey s took place. Within the model, both nets per household and household size were assumed to be normally distributed, with means c_s^h and h_s and standard deviations from the standard errors of the survey data.

In the modeled time series, for each survey s conducted at time t estimated net crop was calculated as an interpolation between values at the beginning and end of the quarter q in which t falls:

$$\Gamma_t = (1 - p)\Gamma_q + p\Gamma_{q+1}$$

where p represents the proportion of q that has elapsed by time t.

Net crop was then calibrated to surveys as:

$$C_s \sim N(\Gamma_t, \sigma_s^C) \quad \text{T}(C_x - 3\sigma_s^C, C_s + 3\sigma_s^C)$$

where the truncation constrains model estimates to hew closely to survey reports, which are highly reliable and widely used.

After calibration, estimates of net crop Γ_q^{LLIN} and Γ_q^{cITN} were summed to generate a time series of total ITN crop, Γ_q^{tot}. This is converted to per-capita ITNs using $\gamma_q^p = \frac{\Gamma_q^{tot}}{P_q}$. This is the main output of the stock and flow model, which is passed on

to calculate household-size-specific indicators ρ_h^0 and μ_h to find a time series of ITN access by country.

Conversion to Access

Symbol	Meaning	Units
ρ_h	Proportion of households of size h with no nets	Proportion
μ_h	Mean net count given ownership of at least one net, among households of size h	Nets
Λ^{acc}	National-level ITN access	Proportion of population with access

Table 3.5: Parameter definitions for conversion of net crop to net access.

The simplest way to convert from ITN crop Γ to ITN access Λ^{acc} would be $\Lambda^{acc} = min(\frac{2\Gamma}{P}, 1)$, using the definition of access as ownership of two nets per person. However, this simple equation does not account for the overdispersion of nets to certain households, a well-documented phenomenon explored in depth in Kilian et al. [40] and Bhatt et al. [12]. One of the most widely available predictors of sufficient access is household size, with nationally-representative surveys consistently showing that access to nets decreases with household size. Therefore, while converting from crop to access, country-specific household size distributions must be taken into account. As discussed in Bhatt et al. [12], the distribution of access among household sizes is well-represented by a zero-truncated Poisson distribution, requiring the estimation of both the proportion ρ_h of households with no nets and the mean net count μ_h among households with at least one net.

Following Bhatt et al. [12], this analysis defined a bivariate joint probability function $H = \mathbb{P}(h, n) \in \mathbb{R}^2$ as the probability mass of a household of size h owning n ITNs. For a given h, the conditional probability distribution $\mathbb{P}(n|h)$ was estimated

as a zero-truncated Poisson distribution [41] characterized by ρ_h and μ_h. Therefore,

$$\mathbb{P}(n = 0|h) = p_h\rho_h$$
$$\mathbb{P}(n > 0|h) = p_h(1 - \rho_h)\frac{(\mu_h)^n}{n!(e^{\mu_h} - 1)}$$

with p_h the proportion of all households having size h. The distribution of household sizes p_h was calculated on a per-survey basis. H is a 10x40 matrix with rows representing household size h and columns representing net count n. These dimensions were chosen because some surveys only collect household size information up to ten people, and 40 is a sufficiently large net count to characterize all meaningful features of these distributions. The entries of H were renormalized as $H_{hn} = \frac{hH_{hn}}{\sum_{h,n} hH_{hn}}$ to convert them from probabilities among households to probabilities among the population. Using the definition of access as "the proportion of people able to sleep under an ITN, assuming one ITN per two people", national access among households of size h with n nets is then:

$$\Lambda^{acc}_{h,n} = \min(\frac{2n}{h}, 1)H_{hn}$$

And overall national access is the sum

$$\Lambda^{acc} = \sum_{h,n} \Lambda^{acc}_{h,n}.$$

Parameters ρ_h and μ_h were defined as polynomial functions of h and the national

net crop per capita, γ^c:

$$\begin{aligned}
\beta_0 &\sim \text{Unif}(-50, 50) \\
\beta_1 &\sim \text{Unif}(-3, 3) \\
\beta_2 &\sim \text{Unif}(-1, 1) \\
\beta_3 &\sim \text{Unif}(-100, 100) \\
\beta_4 &\sim \text{Unif}(-300, 300) \\
\beta_5 &\sim \text{Unif}(-300, 300) \\
\tau &\sim \Gamma(0.1, 0.1) \\
\nu_h &= \beta_0 + \beta_1 h + \beta_2 h + \beta_3 \gamma^c + \beta_4 {\gamma_c}^2 + \beta_5 {\gamma_c}^3 \\
\text{Empirical Logit}(\rho_h) &\sim N(\nu_h, \tau)
\end{aligned}$$

And:

$$\begin{aligned}
\beta_h &\sim \text{Unif}(-20, 20) \\
\tau &\sim \Gamma(0.1, 0.1) \\
\mu_h &\sim N(\beta_h, \tau)
\end{aligned}$$

See Bhatt et al. [12] for validation of these parameterizations. Equation coefficients were fit in Stan using `rstan` version 2.19.3 with all available household-level surveys. It should be noted that the mean of a zero-truncated Poisson with rate parameter λ is $\frac{\lambda}{(1-e^{-\lambda})}$ rather than simply λ. To coerce the mean of $\mathbb{P}$ to equal μ_h for all $n > 0$, the solution of $\mu_h = \frac{x}{(1-e^{-x})}$ was used rather than μ_h.

Quarterly values of ρ_h^0 and μ_h were predicted using these coefficients and the estimated net crop values γ^c from the stock and flow model, with uncertainty propagated at the draw level. These quarterly estimates were interpolated to the monthly time scale, and the monthly estimates used to calculate access as defined above. Thus,

the final result of this modeling process was a monthly time series from 2000 through 2020 of ITN access at the national level for the 40 countries of interest.

3.3.4 Sensitivity Analysis

Each of the 40 countries included in the stock and flow model have between one and ten nationally-representative surveys available for use in model fitting (Table 3.2). To assess sensitivity of this model framework to both survey count and individual survey values, a series of tests were run on the eight countries with more than five survey points.

Country	**Survey Count**
Ghana	6
Rwanda	6
Sierra Leone	6
Tanzania	6
Zambia	8
Malawi	9
Nigeria	9
Senegal	10

Table 3.6: Countries included in sensitivity analysis.

Sensitivity analyses included all available manufacturer and NMCP data, since the completeness of these data types does not change meaningfully across countries. Only sensitivity to survey data was tested.

Consider country c, with N_c available surveys. The model was rerun $N_c - 1$ times, starting from a single survey point and adding an additional survey each round, to test the sensitivity of model estimates to the number of surveys included. To test the importance of the order of survey inclusion, each sequence of $1 : N_c - 1$ models was run three different ways: in chronological order, in reverse chronological order, and in random order. The "random order" option shows only one possible permutation of non-chronological sequence, but it still gives some intuition into the

impact of survey inclusion order on model performance.

The main performance metric was out-of-sample LLIN crop Root Mean Squared Error (RMSE): that is, the RMSE of predicted LLIN crop vs observed LLIN crop for the surveys not included in each model run. Comparisons of LLIN median retention time between each held-out model and the full model are also shown.

3.4 Methods: Spatiotemporal Regression

A spatiotemporal regression framework was developed to derive 5km-by-5km pixel-level ITN access and use from the national-level access metrics estimated in the stock and flow model. First, national ITN access was disaggregated spatially, with survey data informing the difference between local and national access. Next, use was estimated from access, with survey data informing the gap between access and use over space and time. Additionally, a process similar to that for access was employed to estimate pixel-level NPC. NPC does not precisely capture ITN coverage, but does capture spatial heterogeneity in ITN distribution and retention that may be useful for crafting policy.

The present framing of this work is an adaptation of Bhatt et al. [1], with the following improvements:

- Cluster-level recalculation of national access to reflect local household size distributions;
- Addition of nets-per-capita as an outcome variable;
- Calculation of relative gain;
- Full propagation of uncertainty from stock and flow model;

- Addition of over two dozen new or previously-overlooked surveys;
- A fully refactored, publicly available codebase with Google Cloud functionality (https://github.com/bertozzivill/map-itn-cube/tree/publication-2021).

3.4.1 Data

National Net Access and Nets Per Capita

From the stock and flow model–see Section 3.3.3.

Household Survey Data

Ninety-five surveys from 28 countries with household-level data were used to fit the geospatial model. Indicators for (i) household size (defined as number of individuals that slept in the surveyed households the night prior to the survey), (ii) number of ITNs, and (iii) number of individuals that slept under an ITN the night prior to the survey were collected. The relevant cluster identifications were also collected to allow for aggregation to the cluster level. The final dataset contained 34,352 data points covering 28 countries and 17 years.

Covariates

To model the underlying mean function of the latent Gaussian model, a set of environmental and socioeconomic covariates consisting of rasterised satellite imagery across all of Africa at a 2.5 arc-minute (approximately 5km-by-5km) spatial resolution and monthly, annual, or static temporal resolution (Table 3.7) was collected. All raster covariates were derived from high-resolution satellite images that were gap-filled [42] to eliminate missing data resulting primarily from persistent cloud

cover over equatorial forests. Some covariates are themselves modeled products that capture derived phenomena such as temperature suitability to mosquitoes or human accessibility to cities.

3.4.2 Model

Cluster-Level Estimates of Access, Use, and Nets-Per-Capita

Within a given household, the number of people with access to an ITN was calculated as $acc_h = min(2n_h, pop_h)$, with n_h the number of nets in that household. ITN access within a cluster was calculated as $\lambda^{acc} = \sum_h acc_h / \sum_h pop_h$. The underlying national access from the stock and flow model, Λ^{acc}, was adjusted at the cluster level to reflect variation in household size distribution, $\tilde{\Lambda}^{acc} = \sum_h pop_h \Lambda^{acc} / \sum_h pop_h$.

Cluster-level ITN use was calculated as $\lambda^{use} = \sum_c use_c / \sum_c pop_c$, with use_c the number of people who slept under an ITN the night prior to the survey in a given cluster. Cluster-level NPC was calculated as $\lambda^{npc} = \sum_c itn_c / \sum_n pop_c$, with itn_c the number of nets in a given cluster. Clusters were assigned to a time t based on a weighted average of sampling times. Values of λ^{acc}, λ^{use}, and λ^{npc} that fell within the same geographic pixel were aggregated to a single value.

Access Deviation, Use Gap, and Nets-Per-Capita Deviation

Due to the heterogeneity in net distribution volumes and times within countries and the sparsity of survey data, access is a highly nonstationary process which is challenging to predict at high resolution directly from λ^{acc} (Figure 3.4, top row). To create a stationary process that can readily be mapped, λ^{acc} was detrended by the underlying national mean. The resulting metric, access deviation, is therefore a measure of the subnational variation around an overall country mean (Figure 3.4,

Covariate	**Source**	**Years Available**	**Resolution**
Aridity	WorldClim	-	Static
Elevation	SRTM	-	Static
Slope	SRTM	-	Static
NightTime Lights	VIIRS	2014	Static
Accessibility to Cities	MAP modeled output	2015	Static
Plasmodium falciparum Seasonality	MAP modeled output	-	Static
Potential Evotranspiration (PET)	WorldClim	-	Static
Topographic Moisture Index (TMI)	SRTM	-	Static
Population	MAP modeled output	2000-2020	Annual
Land Cover Class 2: Evergreen Broadleaf Forest	MODIS product	2001-2017	Annual
Land Cover Class 4: Deciduous Broadleaf Forest	MODIS product	2001-2017	Annual
Land Cover Class 5: Mixed Forest	MODIS product	2001-2017	Annual
Land Cover Class 6: Closed Shrublands	MODIS product	2001-2017	Annual
Land Cover Class 7: Open Shrublands	MODIS product	2001-2017	Annual
Land Cover Class 8: Woody Savannas	MODIS product	2001-2017	Annual
Land Cover Class 9: Savannas	MODIS product	2001-2017	Annual
Land Cover Class 10: Grasslands	MODIS product	2001-2017	Annual
Land Cover Class 11: Permanent Wetlands	MODIS product	2001-2017	Annual
Land Cover Class 12: Croplands	MODIS product	2001-2017	Annual
Land Cover Class 14: Cropland/Natural Vegetation Mosaic	MODIS product	2001-2017	Annual
Land Cover Class 16: Barren or Sparsely Populated	MODIS product	2001-2017	Annual
Land Cover Class 17: Water	MODIS product	2001-2017	Annual
Enhanced Vegetation Index (EVI)	MODIS product	2000-2017	Dynamic
Tasseled Cap Wetness (TCW)	MODIS product	2000-2017	Dynamic
Daytime Land Surface Tempreature (LST)	MODIS product	2000-2017	Dynamic
Nighttime Land Surface Tempreature (LST)	MODIS product	2000-2017	Dynamic
Temperature Suitability Index (TSI)	MAP modeled output	2000-2017	Dynamic

Table 3.7: Covariates used in all spatiotemporal models. Abbreviations: Shuttle Radar Topography Mission (SRTM), Visible Infrared Imaging Radiometer Suite (VIIRS), Moderate Resolution Imaging Spectroradiometer (MODIS).

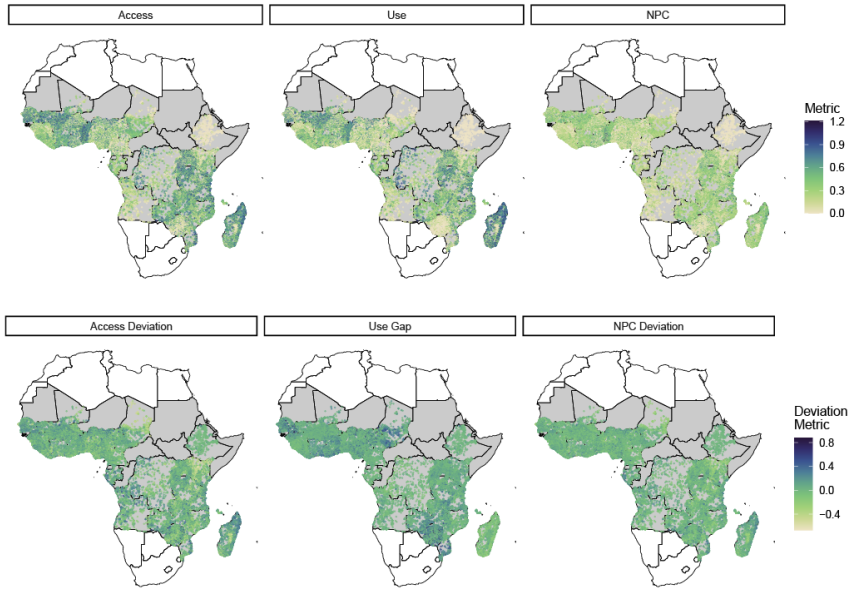


Figure 3.4: Spatial stationarity of deviation metrics (bottom) compared to original metrics (top).

bottom row). Access deviation for a given pixel was calculated as $\lambda^{acc_dev} = \lambda^{acc} - \tilde{\Lambda}^{acc}$. After model fitting and pixel-level prediction of λ^{acc_dev}, pixel-level λ^{acc} was estimated by adding national mean access back onto access deviation. This process is identical for NPC deviation, λ^{npc_dev}.

Net use is similarly spatio-temporally non-stationary, motivating the estimation of the use gap instead. Use gap is simply the pixel-level difference between access and use, $\lambda^{gap} = \lambda^{acc} - \lambda^{use}$.

Both λ^{dev} and λ^{gap} were transformed via the empirical Logit to expand bounds to $[-\infty, \infty]$. This transformation is unnecessary for λ^{npc_dev}, which is not bounded between -1 and 1.

Latent Gaussian Model for Access Deviation, Use Gap, and Nets-Per-Capita Deviation

All outcome metrics were fit using a Bayesian hierarchical model with a Gaussian likelihood and a Gaussian process prior distribution:

$$\theta \sim \pi(\theta)$$
$$y|u, \theta \sim N(Au, \sigma_e^2)$$
$$u|\theta \sim N(\mu, Q^{-1})$$

with $\theta \in \{\sigma_e, \beta, \tau, \kappa, \phi\}$ a vector of hyperparameter priors; $y = \lambda^{acc_dev}$, λ^{gap}, or λ^{npc_dev}; $\mu = \beta X$ a linear basis function for the fixed effects; σ_e^2 the variance for random Gaussian noise; u a Gaussian Markov Random Field (GMRF) representing the spatiotemporal random effect; and A a sparse observation matrix that maps the GMRF to function evaluations at local observations.

The precision matrix Q on the GMRF was defined as a Kronecker product of a temporal process and a spatial process, $Q = Q_t \otimes Q_s$. Temporal correlation was modeled by first order autoregressive dynamics as $w_t = \phi w_{t-1}$ with $w_t \sim N(0, Q_t^{-1})$. The spatial process $w_s \sim N(0, Q_s^{-1})$ is the sparse finite element solution to the stochastic partial differential equation

$$(\kappa^2 - \Delta)^{\alpha/2} \tau u(s) = W(s)$$

where Δ is the Laplacian operator, κ is the spatial scale parameter, α is the spatial smoothness parameter (fixed at $\alpha = 2$), τ controls the variance, and $W(s)$

is the spatial white noise process. Because the spatial process covers a substantial portion of the Earth's surface, s was defined on a spherical manifold in Cartesian $\mathbb{R}^3$. For more details see Lindgren et al. [43].

The models were fit in R using the Integrated Nested Laplace Approximation (INLA, version 20.03.17, https://www.r-inla.org/). To allow for correct posterior coverage and credible intervals under the Laplace approximation, both access and use were transformed via the inverse hyperbolic sine function, $arsinh(x) = ln(x + \sqrt{x^2 + 1})$, thereby allowing the response in the model to be as close to Gaussian as possible. For a given scaling parameter ρ and response $x \in \{\lambda^{acc_dev}, \lambda^{gap}, \lambda^{npc_dev}\}$, with n observations, the inverse hyperbolic sine function was defined as $ihs(x, \rho) = (arsinh(x)\rho)/\rho$. The optimal ρ was found by minimizing the following log-likelihood:

$$l(\rho, x) \propto -nlog(\sum (x - \bar{x})^2) - \sum log(1 + \rho^2 x^2).$$

For more details see Burbridge et al. [44].

3.4.3 Covariate Selection

The full set of covariates (Table 3.7) was selected during a previous iteration of this analysis [1]. The temperature- and aridity- based covariates were selected on the basis of a documented relationship between these variables and net use [16, 45]. The remaining variables were selected either as proxies for socioeconomic status (night-time lights, population, access to cities) or as proxies for other spatially-varying cultural behaviors that may impact net use (land cover types). Because the outputs of this analysis are in turn used as inputs into a regression model to estimate malaria prevalence [37, 46], covariate selection was also constrained to not overlap with the covariates used in the downstream model to avoid circularity.

Three covariate combinations were assessed for predictive performance in the access deviation and use gap models: The full set of available covariates, the full set excluding landcover covariates, and a custom set based on the magnitude and significance of effects in the full model (Table 3.8). Models were compared on the basis of pointwise leave-one-out predictive density $loo_i = log\ p(y_i|y_{-i})$ (LOO) and pointwise Watanabe-Akaike Information Criterion (WAIC).

The model utilizing the full set of covariates performed best by all metrics except for WAIC in the access deviation regression, for which it performed almost identically to the "No Landcover" model (Table 3.9). Based on these results, the full covariate set was utilized in the final model.

3.4.4 Uncertainty Propagation

Both JAGS and INLA return samples from the posterior distributions of the parameters of interest, allowing summary statistics to be explicitly computed as the mean and 95% percentile ranges across these samples. The ideal way to propagate uncertainty would be to run a separate INLA regression for each of the $N_{stockflow}$ stock-and-flow posterior samples, but this was beyond the computational capacity of this analysis. Instead, the mean value across the $N_{stockflow} = 500$ posterior INLA samples was used as the "national mean" from which access deviation and NPC deviation were computed. Then, $N_{geo} = 200$ posterior samples were drawn from each geospatial regression (access deviation, use gap, and NPC deviation). To compute access, use, and NPC at the draw level, 200 samples were selected at random from the stock-and-flow draws and these were added to the INLA samples in place of the stock-and-flow mean values to recapture some stock-and-flow uncertainty in the final result.

Covariate	**Resolution**	**Included in "No Land-cover"**	**Included in Custom Set**
Aridity	Static	Yes	Yes
Elevation	Static	Yes	No
Slope	Static	Yes	No
NightTime Lights	Static	Yes	No
Accessibility to Cities	Static	Yes	No
Plasmodium falciparum Seasonality	Static	Yes	Yes
Potential Evotranspiration (PET)	Static	Yes	No
Topographic Moisture Index (TMI)	Static	Yes	No
Population	Annual	Yes	No
Land Cover Class 2: Evergreen Broadleaf Forest	Annual	No	Yes
Land Cover Class 4: Deciduous Broadleaf Forest	Annual	No	Yes
Land Cover Class 5: Mixed Forest	Annual	No	No
Land Cover Class 6: Closed Shrublands	Annual	No	No
Land Cover Class 7: Open Shrublands	Annual	No	No
Land Cover Class 8: Woody Savannas	Annual	No	No
Land Cover Class 9: Savannas	Annual	No	Yes
Land Cover Class 10: Grasslands	Annual	No	Access Deviation Only
Land Cover Class 11: Permanent Wetlands	Annual	No	Yes
Land Cover Class 12: Croplands	Annual	No	Use Gap Only
Land Cover Class 14: Cropland/Natural Vegetation Mosaic	Annual	No	No
Land Cover Class 16: Barren or Sparsely Populated	Annual	No	Access Deviation Only
Land Cover Class 17: Water	Annual	No	Yes
Topographic Moisture Index (TMI)	Dynamic	Yes	Yes
Tasseled Cap Wetness (TCW)	Dynamic	Yes	Yes
Daytime Land Surface Tempreature (LST)	Dynamic	Yes	No
Nighttime Land Surface Tempreature (LST)	Dynamic	Yes	No
Temperature Suitability Index (TSI)	Dynamic	Yes	Yes

Table 3.8: Covariates used in covariate selection tests.

Model	**Covariate Type**	**LOO**	**WAIC**
Access Deviation	Full Model	**-49,951**	91,705
	No Landcover	-45,951	**91,670**
	Custom	-46,130	92,048
Use Gap	Full Model	**903.6**	**-1902.7**
	No Landcover	888.8	-1872.6
	Custom	859.6	-1814.3

Table 3.9: Validation metrics for covariate sets.

3.4.5 Access vs NPC

To explore allocative inefficiencies in net coverage, country-level, monthly estimates of access and NPC were plotted against each other. A Loess curve between these two variables was also computed and plotted as a visual aid.

3.4.6 Relative Gain

By separately modeling access and use, this analysis was able to distinguish whether lack of net access or lack of net use was the primary driver of low net coverage for those areas that do not achieve universal coverage.

This analysis required the answers to two questions. First, how high could ITN use be if access were kept constant but the use rate were maximized? Second, how high could ITN use be if the use rate were kept constant but access were maximized? These two questions correspond to maps of ITN access and the ITN use rate (use/access) in a given year, respectively. By subtracting true use from each surface, the percentage point gain in coverage from each of these scenarios was calculated. The relative gain of one scenario over another was then assessed by comparing the two gain surfaces.

3.5 Results

3.5.1 Access and Use

Aggregating across all countries in this analysis, after over a decade of steady increases, all ITN coverage metrics have plateaued since 2016 (Figure 3.5, bottom left, all rates are reported among the population at risk of malaria). In 2018 and 2019,

continent-level coverage metrics declined for the first time since mass ITN distributions began, with net crop decreasing from a peak of 360 million (95% credible interval 345-376) in 2017 to 337 million (317-357) in 2019 and access decreasing from 56.3% (54.1-58.8) to 51.0% (48.5-53.6) over the same time period. Mass campaigns in 2020 are estimated to have set a net crop record of 364 million (341-383), but access has continued to plateau at 51.8% (48.8-54.8). The use rate has increased with small fluctuations across the time series examined, from a low of 71.3% (67.6-75.1) in 2004 to a 2020 estimate of 87.1% (83.1-90.3).

Nationally, while most countries have made extensive progress in ITN access and use since 2000, coverage levels remain below the WHO targets. Figure 3.5 shows estimates of ITN access and use at a monthly time scale across all countries in this analysis. The dotted horizontal line represents 80% access or use to indicate the current normative WHO benchmark for universal coverage. Fourteen of 40 countries have ever achieved this coverage, and in 2020 Benin, Mali, Niger, Togo, and Uganda were the only countries estimated to achieve over 80% use (though in some cases lower credible interval bounds were below 80%, and Niger has no survey data to support its estimates). Burundi, Gambia, and Ghana have seen particularly steep coverage declines in recent years, while Mali, Mozambique, Togo, and Uganda have sustained higher coverage. Use aligns closely with access in almost all countries. Madagascar, which has a strong culture of net use despite limited access, shows slightly higher use than access in many years, suggesting the widespread practice of more than two people sharing a net. Ghana, Nigeria, and Zimbabwe experienced a considerable use gap beginning in 2010, but only Ghana has maintained that gap to the present day, with a use rate of 61.8% in the most recent 2019 survey. This corresponds to a use gap of 25.2 percentage points. The modeled estimate of use gap in Ghana for 2019 was 22.8 (19.5-25.8) percentage points, while the aggregated all-country estimates for use rate and use gap were 87.1% (83.3-90.1) and 5.69 percentage points (4.53-7.02), respectively.

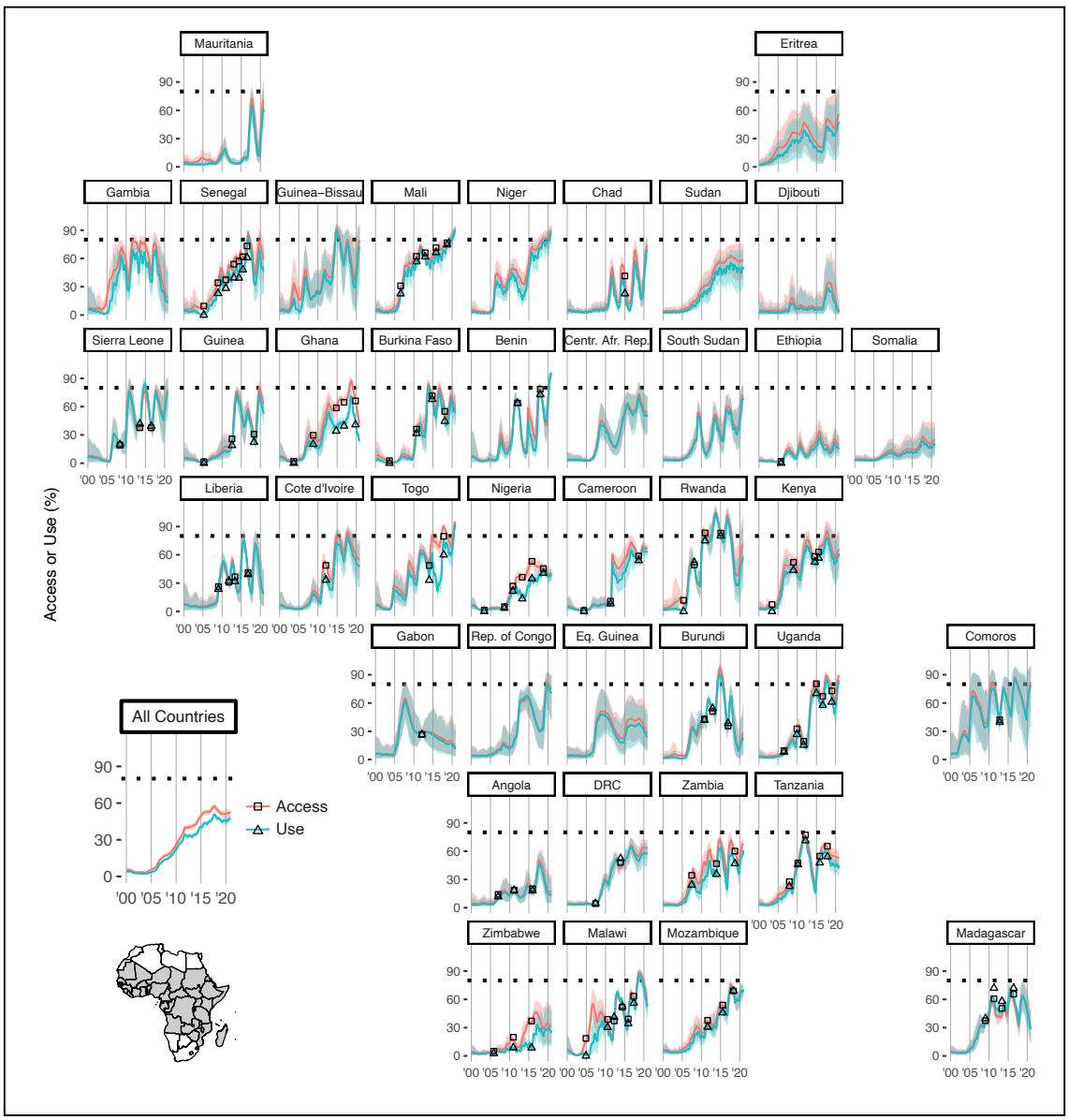


Figure 3.5: National-level insecticide-treated net access and use time series. National-level mean estimates of access (red curves) and use (blue curves) among the population at risk are shown at a monthly time scale. The dotted line indicates 80%, commonly used as a benchmark for universal coverage. Squares and triangles indicate nationally-representative survey values of access and use, respectively. Only surveys that include spatial information are shown here; surveys included in the stock and flow model but not the geospatial model can be found in Table 3.2. Shaded areas represent 95% credible intervals. Spikes in coverage indicate mass net distribution campaigns, with coverage subsequently declining due to attrition. The seasonality of use is evident in the curvature of the blue lines. Centr. Afr. Rep.: Central African Republic; Rep. of Congo: Republic of Congo; DRC: Democratic Republic of the Congo.

A number of studies [16, 17, 45, 47] have established that the ITN use rate often tracks seasonal changes in climate and malaria transmission. While the present analysis was not designed to explore these subannual trends, results show stronger seasonality in use than access even though both models include the same covariates. This finding supports the importance of considering season, climate, and perceived risk when interpreting net use data.

Figure 3.6 highlights subnational heterogeneity in every ITN coverage metric in 2020. Some of this variation is clearly associated with risk-based distribution strategies, such as the stark north-to-south gradients of access and NPC in Mali, Chad, and Sudan. Other subnational heterogeneity may instead suggest areas experiencing challenges to distribution campaigns, such as much of central and northeast Nigeria. Some countries are more homogeneous in their access and NPC estimates, most notably Angola, Zimbabwe, Ethiopia, Somalia, Gabon, Benin, and Togo. Estimated spatial patterns of the ITN use rate differ substantially from those of access and NPC, with higher values overall and notably heterogeneous patterns in Ghana. Western Tanzania stands out as an area of concern, but this may be an artefact of data collection. For both the 2015-16 Demographic and Health Survey and 2017 Malaria Indicator Survey, nets were distributed in western Tanzania over the same time period as survey data collection. Surveys conducted just before large net distributions might explain some of the effect observed in this region.

The uncertainty plots in Figure 3.6 present population-binned mean values and 95% credible interval width for each metric. More saturated colors represent narrower uncertainty, while the pink-to-blue spectrum indicates mean values. For example, in Saharan regions such as Mauritania and northern Mali there is confidence that net access was in the lowest quartile in 2020, while in eastern Tanzania there is moderate certainty that it was in the highest quartile. In the Democratic Republic of the Congo, Zimbabwe, and Kenya, which have not conducted household surveys

in several years, estimates are highly uncertain. Full time-series maps of all metrics and associated uncertainty are available at the Malaria Atlas Project website. This website also hosts an interactive visualization of results and uncertainty, which features an option to plot uncertainty in terms of exceedance probabilities (Figure 3.22) as a complement to the relative uncertainty shown here.

Figure 3.7, Figure 3.8, and Figure 3.9 show national-level time series of the three regression outputs. On the national level, by definition, both access deviation (Figure 3.7) and NPC deviation (Figure 3.8) should be close to zero. However, because national mean estimates (blue horizontal lines) are derived from the stock and flow model for these two measures, the survey data can show nonzero values in country-years where the stock and flow model did not perfectly fit the surveys, and the regression estimates will be similarly nonzero.

Use gap values are not constrained to be near zero. Figure 3.9 also demonstrates the stronger seasonal patterns detected in the use gap regression. While use rate is a derived output that was not directly fitted to data, time series of it are included here for reference (Figure 3.10).

3.5.2 Access vs Nets-Per-Capita

Nets-per-capita and access are highly correlated metrics, but their relationship features important nonlinearities that pose challenges for net allocation (Figure 3.11). When planning mass campaigns, WHO recommends that countries procure one net per 1.8 people at risk (0.56 nets-per-capita) to ensure universal access while accounting for households with an odd number of residents [6, 40]. The tacit assumption is that ITN access will scale linearly with the number of nets distributed. The present analysis demonstrates that this assumption holds at low coverage levels. However, at NPC levels above approximately 0.25 the relationship between NPC and access

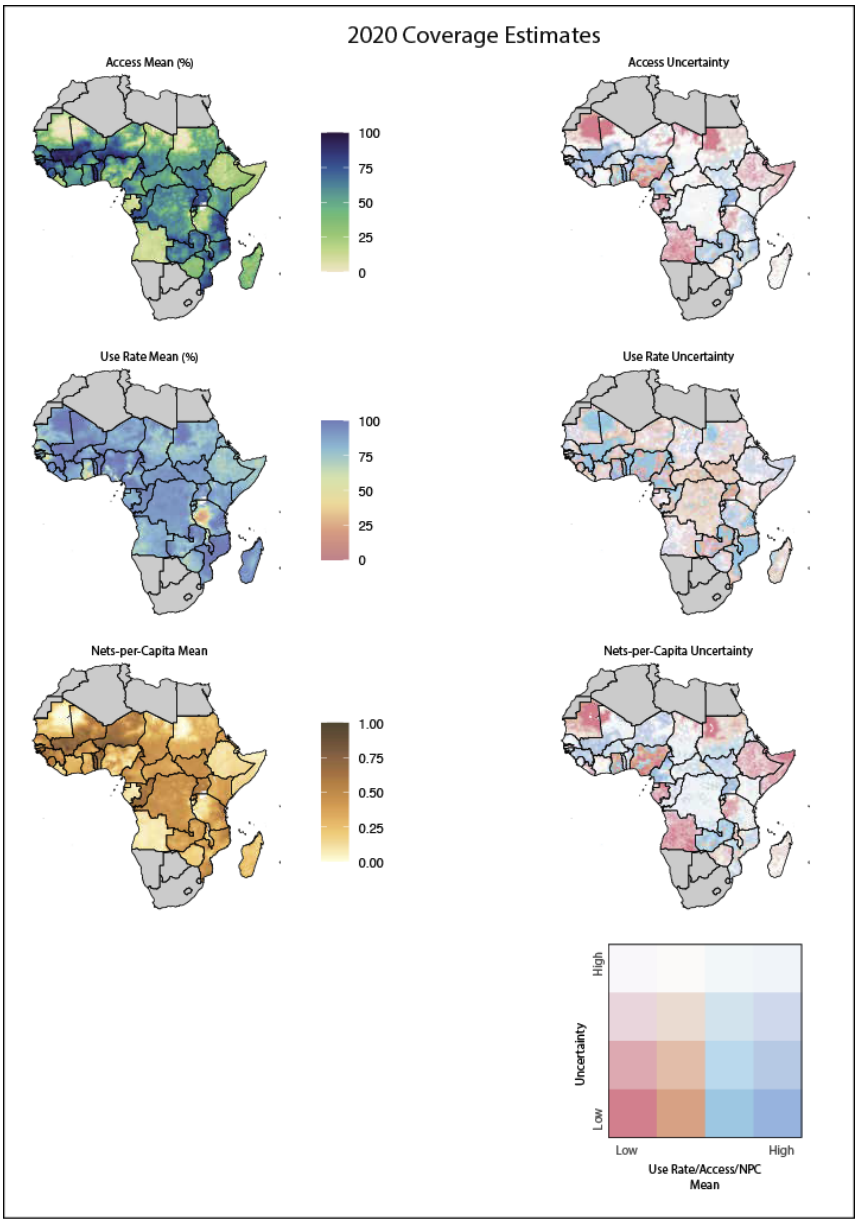


Figure 3.6: Maps of Insecticide-treated net access, use rate, and Nets-Per-Capita (NPC) in 2020, with associated uncertainty. Access is defined as the proportion of the population that could sleep under a net, assuming one net per two people. The use rate is defined as use among those with access. The uncertainty maps use a bivariate scale to convey information about both mean estimates and magnitude of uncertainty. Quantiles of 95% credible interval width are represented by saturation level, with the most uncertain values having the lowest saturation. Quantiles of mean values are represented by hue, with low values in pink and high values in dark blue.



Figure 3.7: Insecticide-Treated Net (ITN) access deviation by country. Solid curves show mean values, and shaded areas represent 95% credible intervals. Survey data mean values shown as solid circles. AFR: Africa. All other abbreviations are ISO3 codes, see Table 3.10 for a key to country names. Solid blue line at zero included as a reference point.

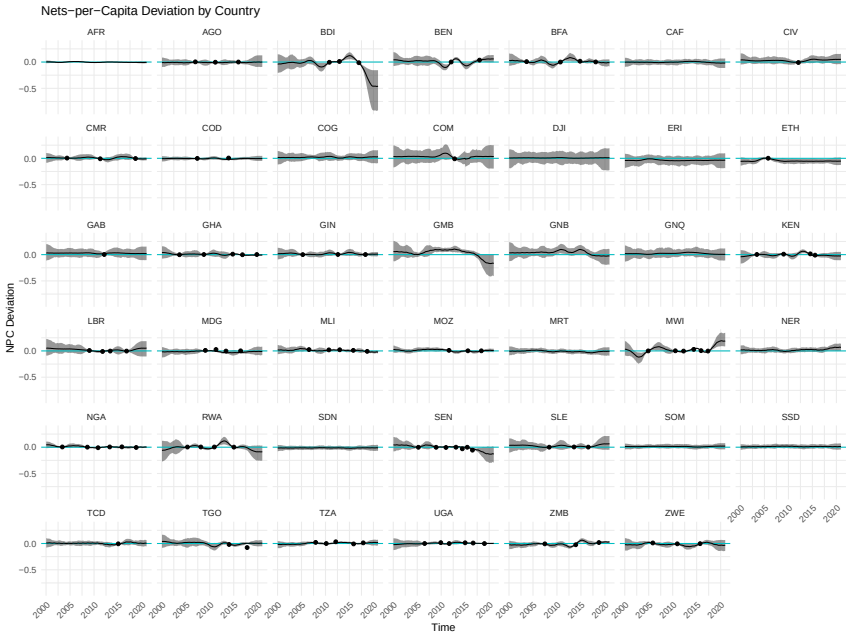


Figure 3.8: Nets-per-capita deviation by country. Solid curves show mean values, and shaded areas represent 95% credible intervals. Survey data mean values shown as solid circles. AFR: Africa. All other abbreviations are ISO3 codes, see Table 3.10 for a key to country names. Solid blue line at zero included as a reference point.



Figure 3.9: Insecticide-Treated Net (ITN) use gap by country. Solid curves show mean values, and shaded areas represent 95% credible intervals. Survey data mean values shown as solid circles. AFR: Africa. All other abbreviations are ISO3 codes, see Table 3.10 for a key to country names. Solid blue line at zero included as a reference point.

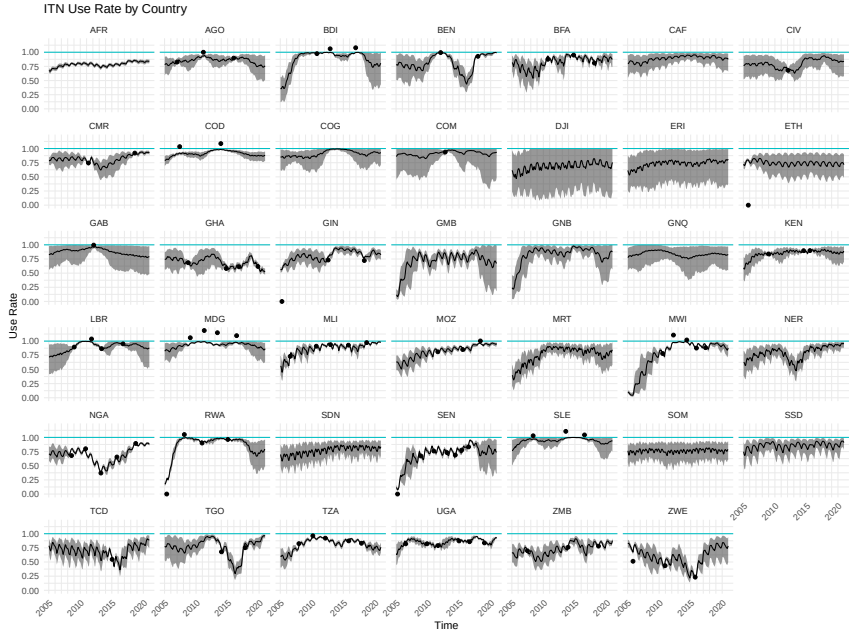


Figure 3.10: Insecticide-Treated Net (ITN) use rate by country. Solid curves show mean values, and shaded areas represent 95% credible intervals. Survey data mean values shown as solid circles. AFR: Africa. All other abbreviations are ISO3 codes, see Table 3.10 for a key to country names. Solid blue line at zero included as a reference point.

begins to plateau, such that attaining coverage levels of 0.5 nets-per-capita corresponds to ITN access well below 0.75, instead of the predicted value of 0.9. This pattern suggests substantial misallocation of nets at higher coverage levels, such that some households are left with insufficient protection despite the total number of nets in a country continuing to increase (Figure 3.11). The results shown here represent modeled estimates, but this relationship is also evident in the survey data alone (Figure 3.12).

Figure 3.11 demonstrates the curvature of the relationship between ITN access and NPC for modeled country-months in the year 2020. Figure 3.12 shows the consistency of this relationship by plotting survey data on the cluster level along with modeled estimates on the country-month level for all years in which survey data were available.

Another way to assess the impact of less-than-perfect net allocation is to calculate a counterfactual "optimal access" metric based on the number of nets-per-capita, $\min(2 * \text{NPC}, 1)$. Plotting this value against estimates for access (Figure 3.13) gives an upper bound of what ITN access could look like with improved allocation.

3.5.3 Net Retention Time

The model generates separate national estimates of median retention time for Long-Lasting Insecticide-Treated Net (LLIN)s and Conventional Insecticide-Treated Net (cITN)s. Because LLINs have been the primary net distributed in Africa for many years, only LLIN retention is reported here. Results estimate shorter LLIN lifespans than the three-year retention time assumed on average by WHO and other policy-making bodies (Figure 3.14, Table 3.10) [48]. Of the 40 countries in this analysis, 35 show LLIN median retention times of under three years, with an overall median value of 1.64 years (IQR 1.33-2.37). Many of the countries with the most extreme retention

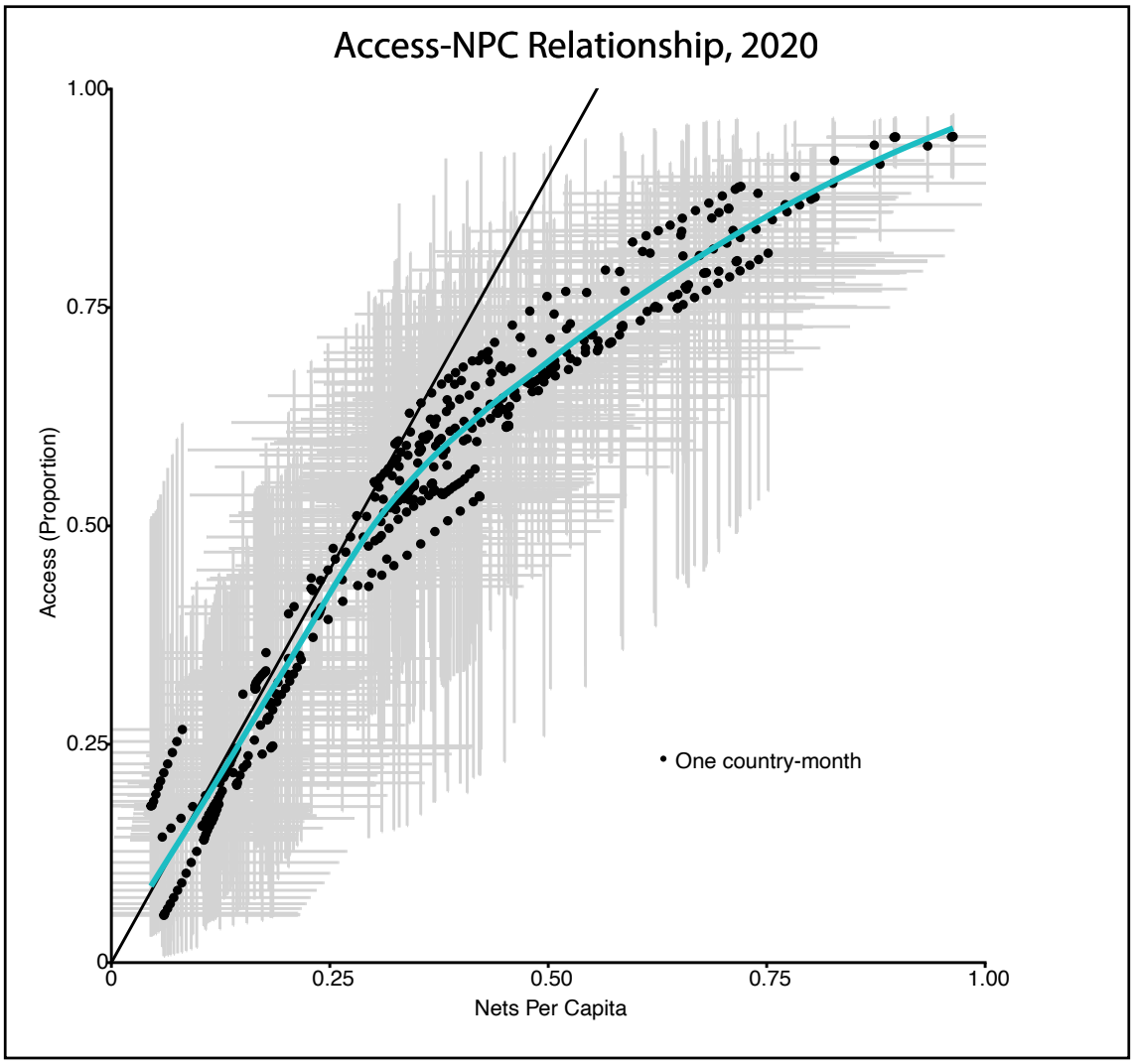


Figure 3.11: Relationship between access and Nets-Per-Capita (NPC) in 2020. Each dot represents a modeled country-month, with black dots indicating mean values and gray bars indicating 95% credible intervals for each metric. The solid line has a slope of 1.8, showing the relationship between NPC and access presumed by population-based net procurement decisions. The blue curve shows a Loess fit through the estimated points. The expected linear relationship between NPC and access holds well at low coverage levels, but the true relationship tapers off at NPC values greater than 0.25 and access over 0.5. This plateauing of access despite high numbers of nets distributed per capita suggests inefficiencies or redundancies in net distribution at these coverage levels, such that those who should be receiving nets are still left without access.

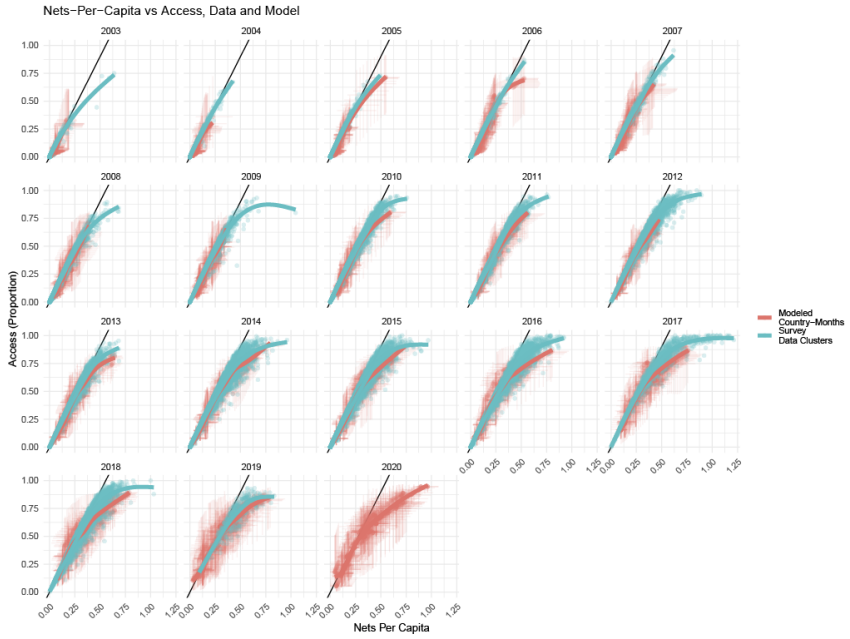


Figure 3.12: Comparison of modeled (red) and raw data (blue) relationship between nets-per-capita and Insecticide-Treated Net (ITN) access.

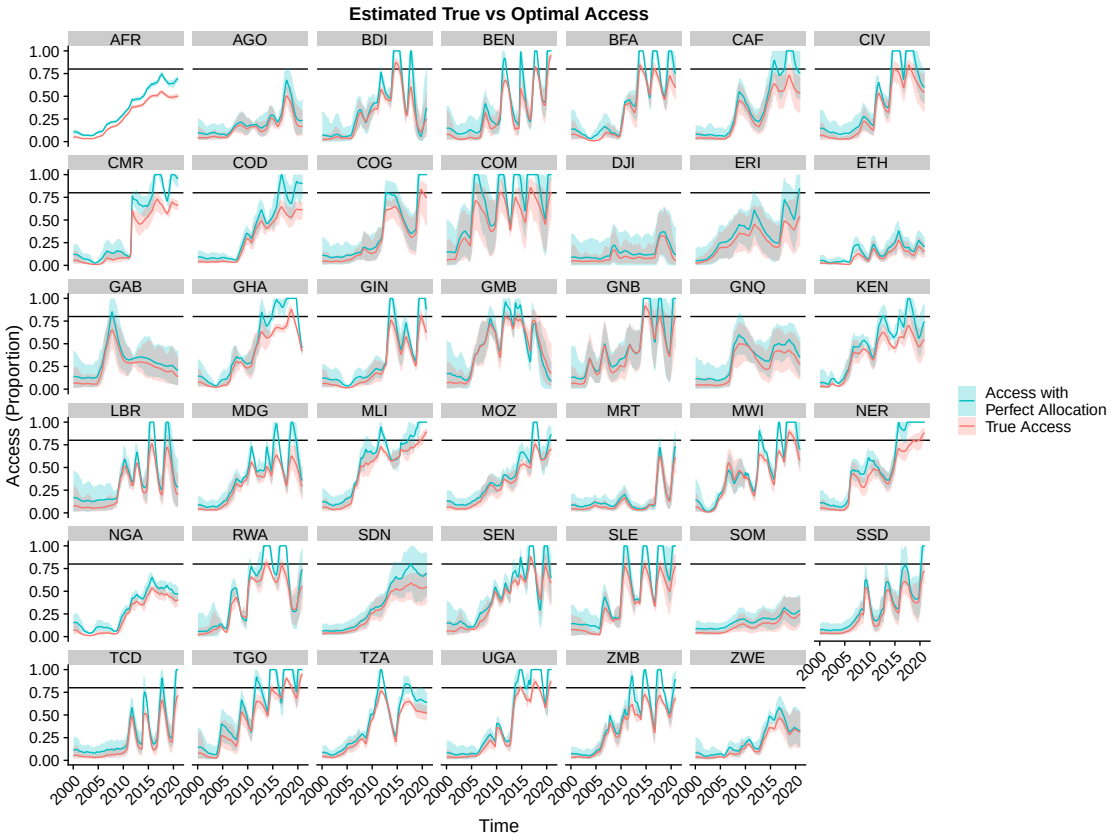


Figure 3.13: Insecticide-Treated Net (ITN) estimated true vs optimal access by country. Estimated true access shown in red, and hypothetical access with perfect allocation shown in blue. Solid curves show mean values, and shaded areas represent 95% credible intervals.

times are also those with the fewest surveys available for fitting, and thus these values may not be reliable (see sensitivity analysis in Section 3.5.6). Among countries with stable model fits, Cameroon is the only country with a median retention time over three years (3.49 years, 95% CI 3.24-3.78). Thirteen countries have median retention times whose mean and upper credible interval bound are below 1.5 years, from Mozambique (1.34 years, 1.21-1.50) to South Sudan (1.02 years, 1.01-1.04). The model prior distribution on retention time was bounded at one year, suggesting that some countries might show even shorter median retention times in an unbounded setting. However, given the small number of surveys in these countries, such results might indicate underspecified models rather than truly short retention times. See Section 3.7.1 for caterpillar plots of the retention time parameter for all countries.

LLIN Median Retention Times.

Country	**ISO3**	**Median Retention Time in Years**	**Survey Count**
Angola	AGO	1.10 (1.01-1.26)	3
Benin	BEN	1.07 (1.01-1.17)	4
Burkina Faso	BFA	1.58 (1.41-1.76)	5
Burundi	BDI	1.31 (1.14-1.47)	4
Cameroon	CMR	3.49 (3.24-3.78)	4
Central African Republic	CAF	1.90 (1.56-2.26)	2
Chad	TCD	1.03 (1.01-1.08)	1
Comoros	COM	2.13 (1.81-2.39)	1
Congo (Republic of)	COG	2.91 (2.31-3.65)	3
Cote d'Ivoire	CIV	1.69 (1.51-1.86)	4
Democratic Republic of the Congo	COD	1.41 (1.15-1.64)	3
Djibouti	DJI	1.05 (1.01-1.13)	2
Equatorial Guinea	GNQ	3.59 (3.27-3.79)	1
Eritrea	ERI	3.01 (1.94-3.79)	1
Ethiopia	ETH	1.33 (1.19-1.48)	4
Gabon	GAB	3.34 (2.63-3.79)	1
Gambia	GMB	1.62 (1.39-1.85)	4
Ghana	GHA	1.78 (1.67-1.90)	7
Guinea	GIN	1.51 (1.28-1.75)	4

Continued on next page

LLIN Median Retention Times, cont.			
Country	**ISO3**	**Median Retention Time in Years**	**Survey Count**
Guinea-Bissau	GNB	1.38 (1.01-2.16)	2
Kenya	KEN	2.26 (1.98-2.58)	5
Liberia	LBR	1.03 (1.01-1.07)	4
Madagascar	MDG	1.65 (1.48-1.81)	4
Mali	MLI	2.81 (2.46-3.14)	6
Mauritania	MRT	1.07 (1.01-1.16)	2
Mozambique	MOZ	1.34 (1.21-1.50)	3
Malawi	MWI	1.33 (1.21-1.45)	9
Niger	NER	3.50 (3.25-3.78)	2
Nigeria	NGA	2.22 (2.00-2.47)	9
Rwanda	RWA	1.59 (1.48-1.70)	6
Senegal	SEN	1.35 (1.22-1.48)	10
Sierra Leone	SLE	1.47 (1.31-1.63)	5
Somalia	SOM	2.35 (1.02-3.66)	1
South Sudan	SSD	1.02 (1.01-1.04)	3
Sudan	SDN	2.91 (2.11-3.77)	2
Tanzania	TZA	2.15 (1.88-2.43)	6
Togo	TGO	2.42 (2.21-2.61)	4
Uganda	UGA	1.66 (1.55-1.78)	6
Zambia	ZMB	1.32 (1.21-1.44)	9
Zimbabwe	ZWE	2.79 (2.26-3.38)	5

Table 3.10: Median retention times and 95% CI's for LLINs from the stock and flow model. Number of surveys used to fit each country model also listed.

3.5.4 Relative Gain: Optimising Access or Use?

The relative gain analysis supports other evidence that lack of access is the primary barrier to universal coverage. The top row of Figure 3.15 shows the estimated use map for 2020. The second row shows what the use map would look like if access were held constant and the use rate were 100% (left), versus if the use rate were held constant but access were 100% (right). The bottom row represents these two plots in terms of the percentage point increase in use that could be achieved by increasing the use rate as opposed to increasing access.

In 2020, with the exception of western Tanzania and scattered areas of the Sahel,

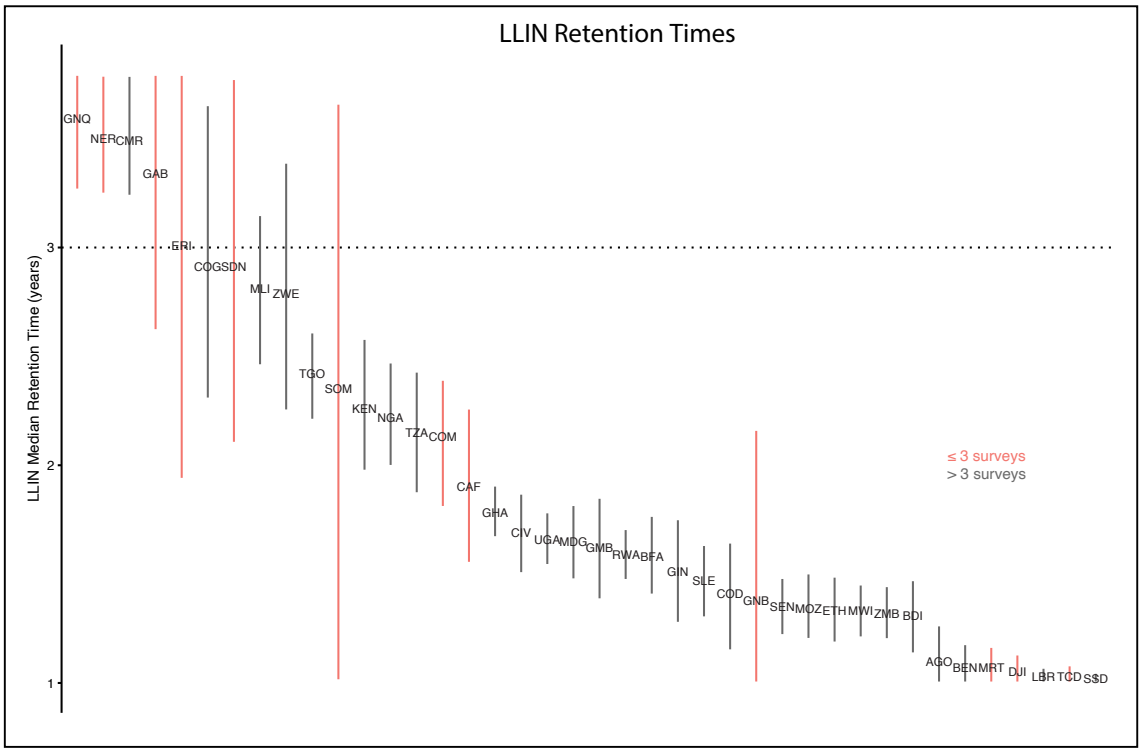


Figure 3.14: Long-Lasting Insecticide-Treated Net (LLIN) median retention times. Stock-and-flow estimates of median LLIN retention time by country, ordered from highest to lowest. Countries are labeled by ISO3 code. Country labels are positioned at mean parameter values, while vertical bars indicate 95% credible interval width. Countries with fewer surveys have less stable model fits; those having fewer than three surveys are indicated in red. The lower bound of LLIN retention time was capped at one year during model fitting. Table 3.10 shows all data for this figure in numerical format and maps ISO3 codes to country names.

lack of access to a net was a far greater barrier to coverage than deficiencies in use. While the overall impact of increasing access is greater than the impact of increasing use rate for all years, the spatial patterning of this relative gain varies substantially from year to year. For example, in 2015 it is not Tanzania but Zimbabwe and Chad where use rate is a larger barrier than access (Figure 3.16). The season in which surveys are conducted may play a substantial role in the volatility of use rate estimates [14].

3.5.5 Stock and Flow Model Crop-to-Access Regression Coefficients

Shown in Table 3.11 and Table 3.12 are the outputs from the polynomial regression to determine coefficients for the crop-to-access regression.

Variable	**Mean (Lower 95%-Upper 95%)**
β_0	4.07 (3.94-4.19)
β_1	-0.286 (-0.327- -0.245)
β_2	0.0168 (0.0131-0.0205)
β_3	-24.8 (-25.9- -23.6)
β_4	45.7 (41.2-50.1)
β_5	-31.6 (-36.6- -26.6)
τ	0.406 (0.388-0.425)

Table 3.11: Regression coefficients for determining the proportion of households with no nets, ρ_h.

3.5.6 Stock and Flow Sensitivity Analysis

Figure 3.17 shows the out-of-sample RMSE (in units of millions of LLINs) for each country-iteration of the analysis. When surveys are added in chronological order, the RMSE trend is nonlinear, often peaking at some intermediate number of surveys before declining. Adding surveys in reverse chronological order shows a more stable trend, with RMSE generally declining with each additional survey. One major

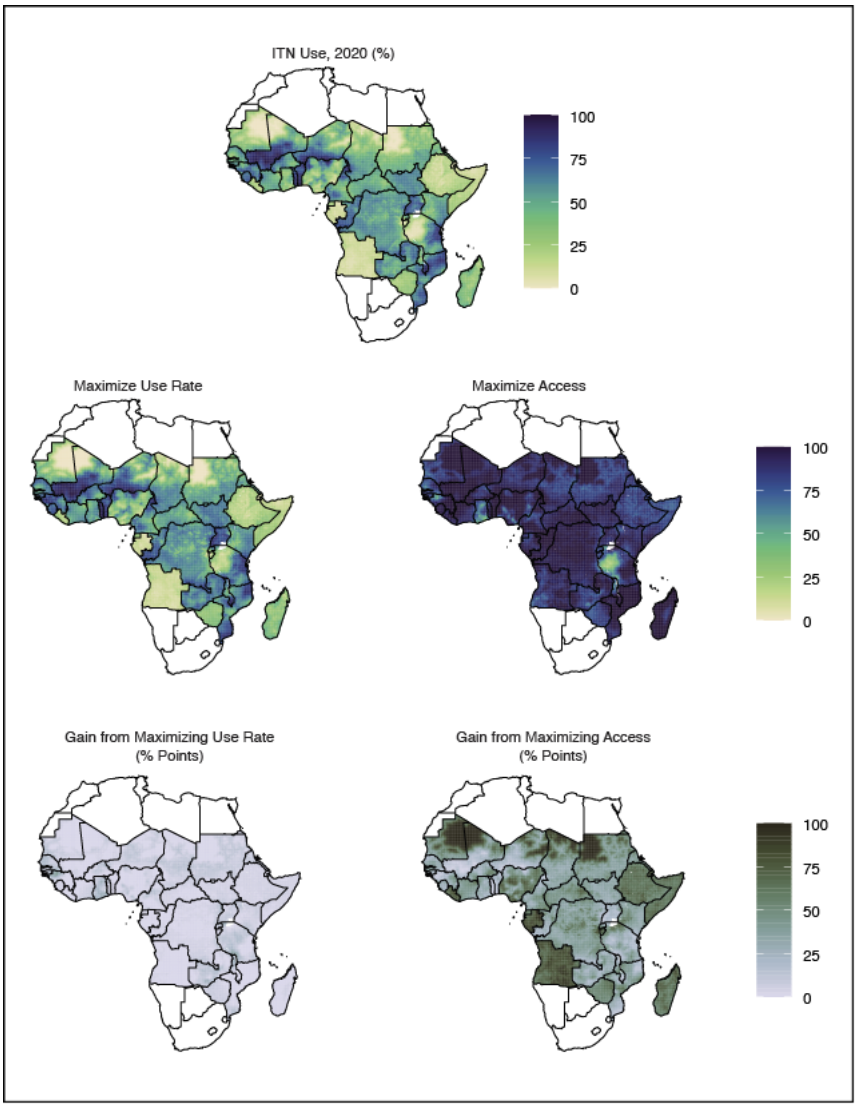


Figure 3.15: Magnitude of change in Insecticide-Treated Net (ITN) use possible from increasing use rate versus increasing access. The top row shows estimated ITN use in 2020. The second row shows what use could be if access remained unchanged and the use rate were set to 100% (left), compared to if the use rate remained unchanged and access were set to 100% (right). The final row shows the magnitude gain in use from each of these two scenarios. With few exceptions, increasing access has a larger impact than increasing the use rate.

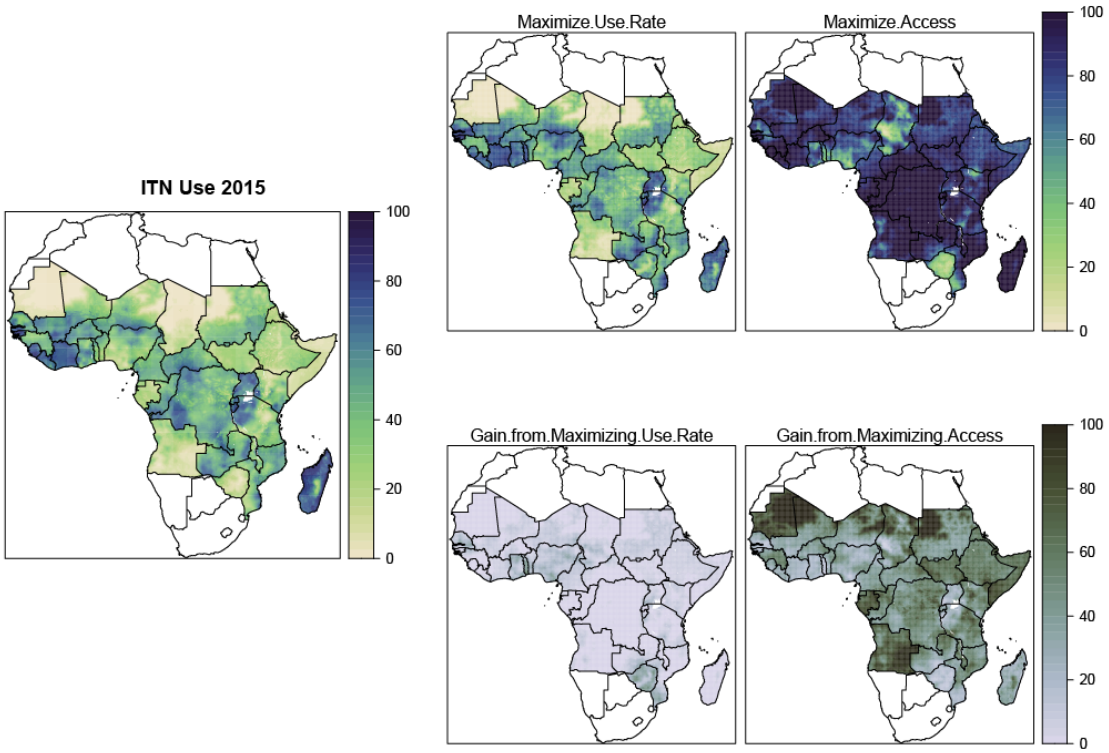


Figure 3.16: Relative gain example from 2015. The leftmost plot shows estimated use, the top two plots on the right show use if the use rate were maximized versus if access were maximized, and the bottom two plots on the right show the percentage point increase in use for these two scenarios.

Household Size	**Intercept**	**Slope (β_h)**
1	0.15 (0.01-0.3)	1.03 (0.5-1.56)
2	0.24 (0.08-0.39)	2.08 (1.54-2.63)
3	0.41 (0.26-0.56)	2.48 (1.94-3.02)
4	0.52 (0.37-0.67)	3.22 (2.68-3.77)
5	0.65 (0.5-0.8)	3.88 (3.35-4.41)
6	0.76 (0.6-0.91)	4.38 (3.82-4.92)
7	0.81 (0.66-0.96)	5.02 (4.49-5.55)
8	0.83 (0.68-0.98)	5.72 (5.18-6.27)
9	0.87 (0.71-1.02)	6.3 (5.76-6.84)
10	1.06 (0.91-1.21)	7.89 (7.35-8.42)

Table 3.12: Regression coefficients for determining mean nets per household, μ_h.

exception is Sierra Leone, which has a large peak in RMSE at three surveys. The random permutation approach splits the difference between these two, but generally behaves slightly more like the "reverse chronological" than the "chronological". For a given country-survey count, chronological RMSE is usually higher than reverse-chronological RMSE. Including only the low-net-count surveys early in the time series yields the highest RMSEs. Otherwise, there is not a consistent marker of which survey will have the highest impact on model fit. In Sierra Leone, for example, all models take a step change in RMSE upon the inclusion of the middle-of-the-series 2011 survey, while in Tanzania it is the two surveys at the end of the time series that most impact RMSE. For countries like Senegal, no one survey has an outsized impact.

Figure 3.18 shows the evolution of estimated LLIN median retention time (also called half-life). The solid line through each plot is the half-life estimated by the full model. Again here, with the exception of Sierra Leone, the "reverse chronological" sequence converges quickly to the value estimated by the full model, whereas the chronological sequence varies considerably before arriving at the full model's result. When just one survey is included, half-lives are more likely to be overestimates than underestimates of the full model value.

These results suggest that stock and flow time series fit to a small number of surveys, especially if those surveys occur early in the time series, may not be reliable estimators for LLIN half-life or net time series in the country they represent. As shown in Figure 3.14, countries with three or fewer surveys are likely to fall at the extreme ends of the LLIN half-life spectrum. In these countries, local knowledge and additional data collection will be crucial to understanding the true circumstances of net distribution, retention, and use. See Section 3.7.2 for full plots by country of sensitivity analysis results.

3.5.7 Geospatial Regression Coefficients and Model Validation

Figure 3.19 shows coefficient values for all model covariates. Note that these values are in transformed variable space, not level space.

Figure 3.20 plots true vs predicted values for all survey data points. Posterior coverage was evaluated by leave-one-out Probability Integral Transform (PIT). For well-calibrated posterior distributions, the distribution of PITs should be close to Uniform. PIT distributions showed acceptable posterior coverage for all models, though the NPC regression will require more adjustment in future iterations to ensure proper calibration (Figure 3.21).

3.5.8 Example of Exceedance Surfaces for Uncertainty Visualization

This analysis communicates geospatial uncertainty using an uncertainty quantile approach. Another way of visualizing this uncertainty is through an exceedance approach. For Bayesian models such as these, exceedance surfaces demonstrate

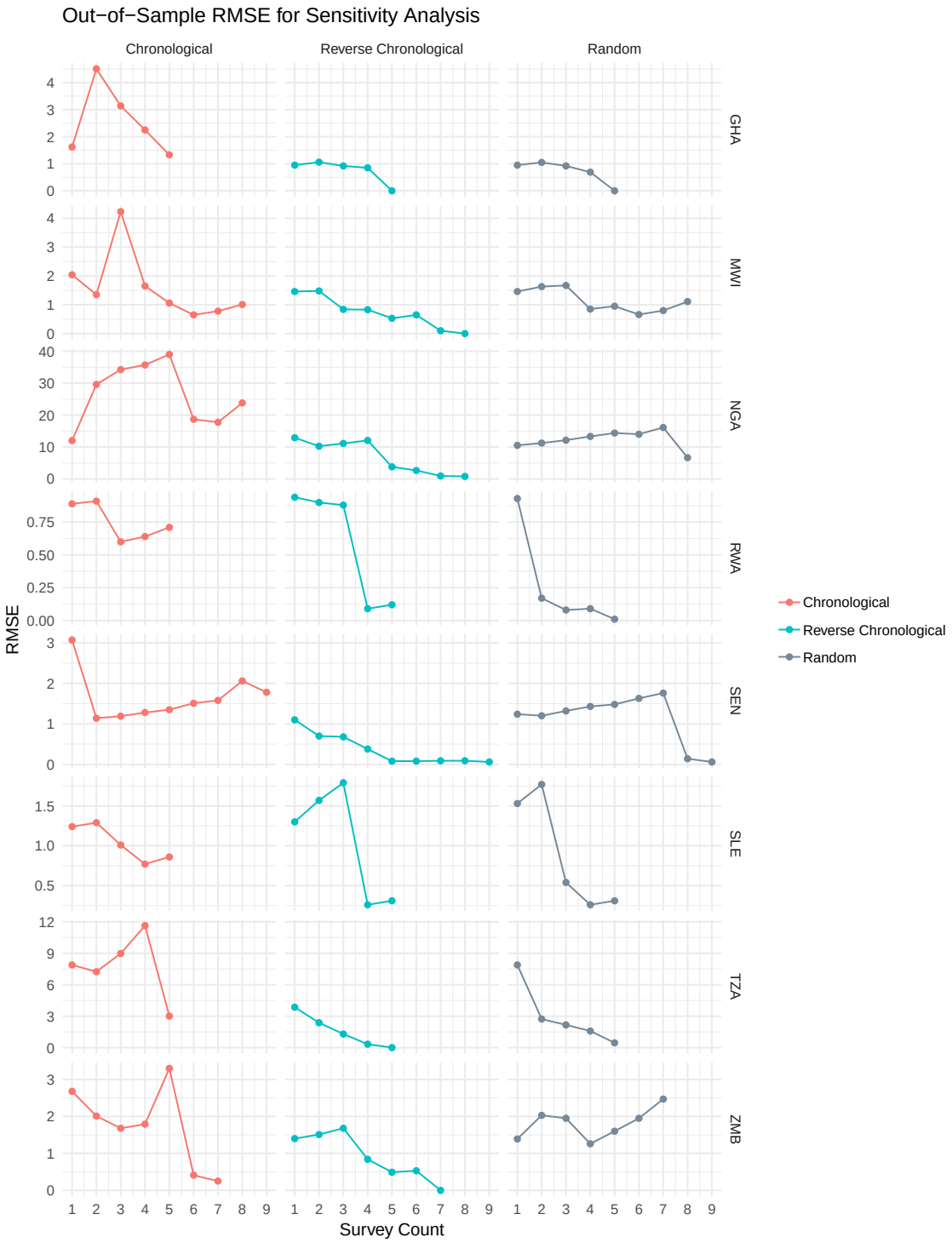


Figure 3.17: Out-of-sample root mean squared error evolution in sensitivity analysis (Units: millions of LLINs)

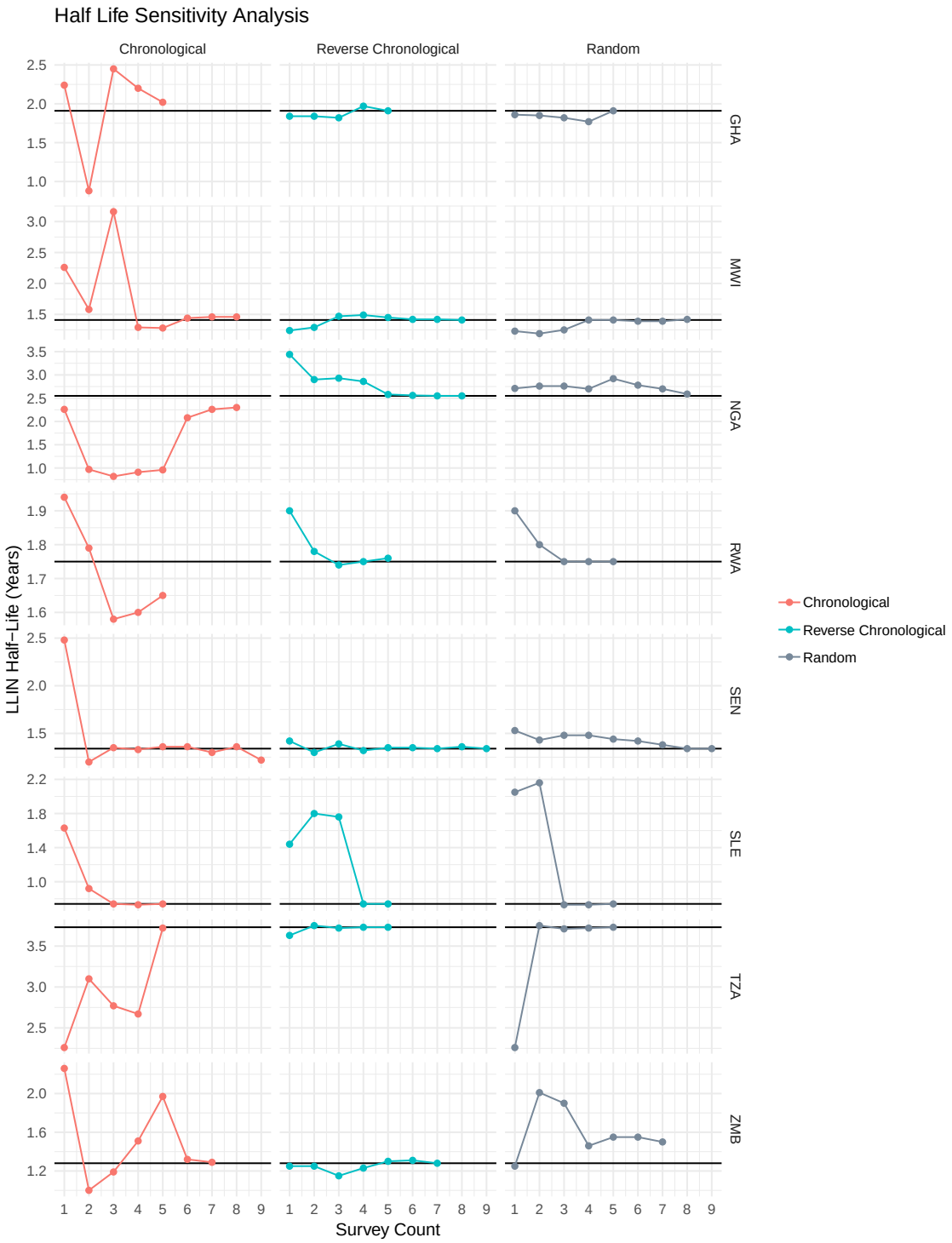


Figure 3.18: Median retention time (or "half life") evolution in sensitivity analysis.

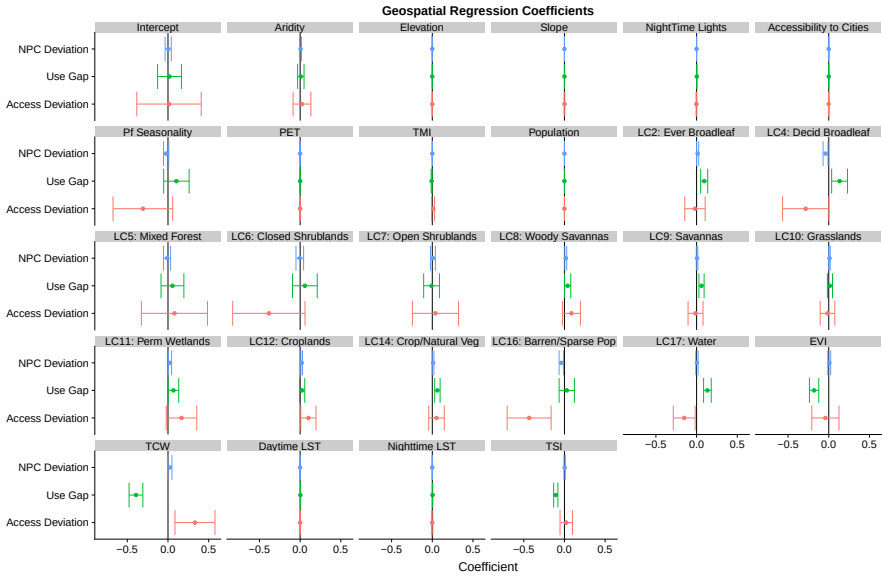


Figure 3.19: Geospatial regression coefficients. Dots represent mean values, and error bars represent 95% credible intervals. For abbreviation definitions see Table 3.7.

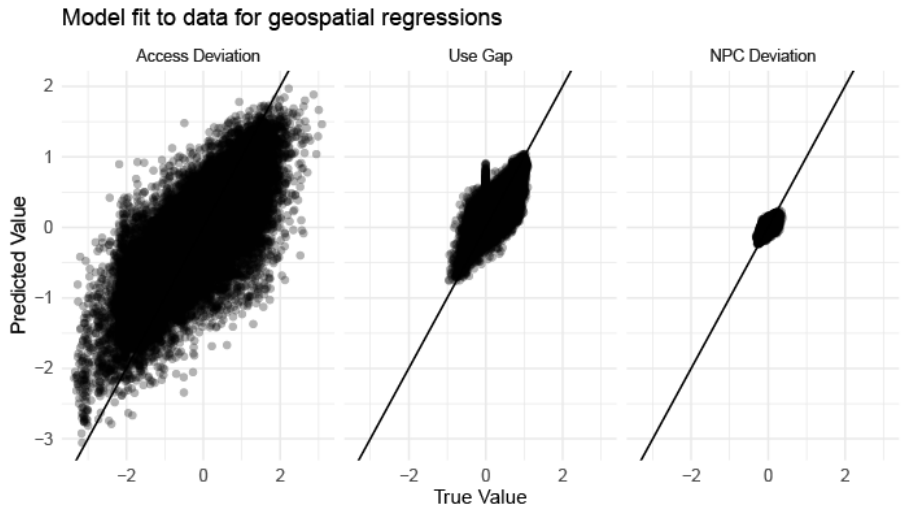


Figure 3.20: Geospatial regression model fits.

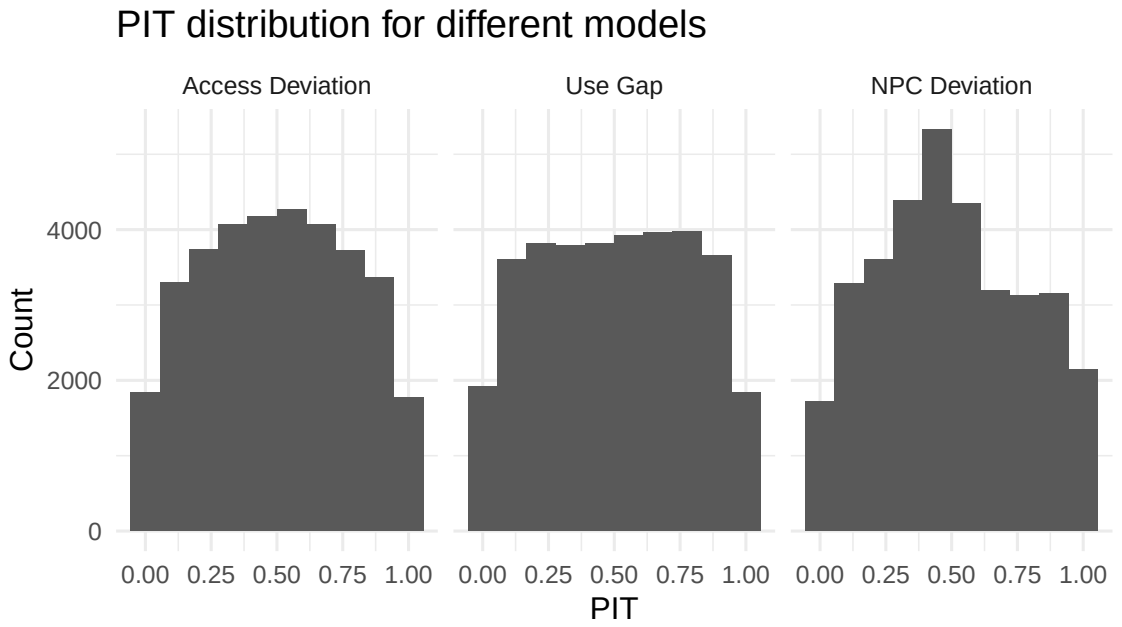


Figure 3.21: Geospatial regression Probability Integral Transform (PIT) distributions.

the proportion of posterior draws that are above or below a given threshold. As an example, Figure 3.22 shows posterior mean maps (central column) along with several positive and negative exceedance surfaces. The top row demonstrates that all 200 posterior draws showed an access estimate below 0.3 in Angola in 2020, whereas in Botswana only a fraction of draws were below this threshold. In contrast, almost all of Botswana had posterior draws below 0.5 in this same year. An interactive visualization to explore a range of cutoffs is available at the Malaria Atlas Project website.

3.6 Discussion

The goal of this analysis was to fully characterize ITN access, use, and nets-per-capita over space and time for 40 sub-Saharan African countries. Results showed that ITN distributions have increased enormously since the year 2000 and those who own nets tend to use them, but a combination of insufficient net volume, distribution

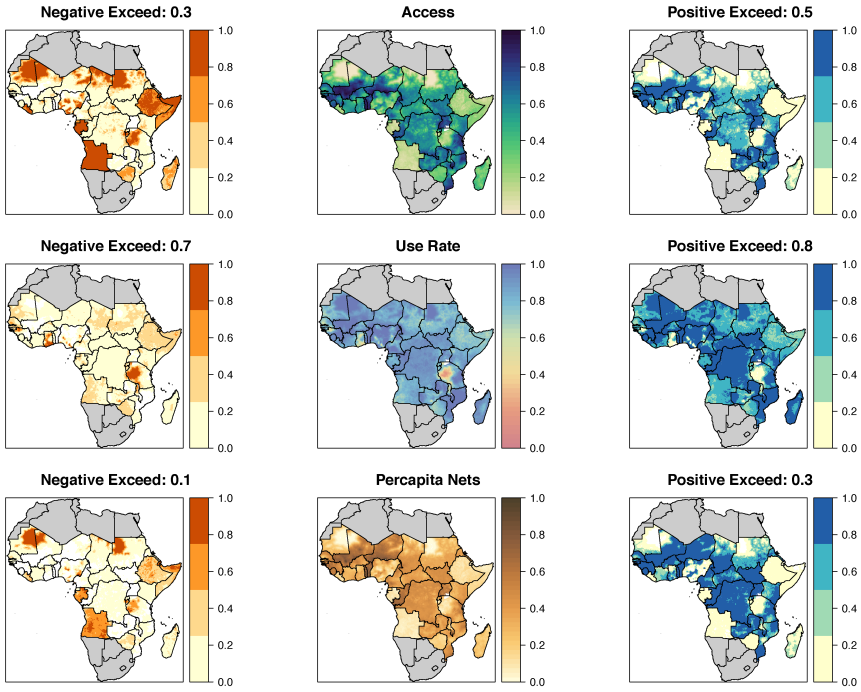


Figure 3.22: Example 2020 exceedance surfaces. See https://malariaatlas.org/research-project/metrics-of-insecticide-treated-nets-distribution/exceedance-surfaces/ for an interactive exceedance explorer for all metrics and years.

inefficiency, and short retention times keep ITN access and use below WHO targets of 80% coverage.

Despite prodigious effort on the part of national malaria programs, mass net distributions have rarely been able to distribute sufficient nets to all those in need to attain universal coverage. Historically large campaigns completed in 2020 despite COVID-related disruptions pushed several countries over the 80% threshold for access and use. These include Benin, Mali, Niger, Togo, and Uganda. If current trends continue, some countries, such as Sierra Leone, Burkina Faso, Kenya, Malawi, and Mozambique, may distribute sufficient nets for universal coverage within one or two mass campaigns, whereas other countries such as Cote d'Ivoire, Liberia, and Tanzania have lost ground or stalled progress. However, official policy is now shifting away from mandates of universal ITN coverage and towards recommendations for more holistic and locally-tailored strategies that utilize a range of tools to provide

universal malaria protection [11, 49]. This would reduce the relevance of the 80% target in favor of a more nuanced national and subnational approach.

This analysis adds to the evidence [4, 12] of net misallocation at higher coverage levels, leading to lower-than-expected ITN access for a given number of nets-per-capita. Some amount of misallocation is likely unavoidable as a corollary to well-established econometric relationships that show dramatically increasing intervention costs at high coverage levels [50], but considerations of how to mitigate this effect are still worthwhile. One known source of net misallocation is the mass campaign practice of capping the number of nets any single household can obtain, disadvantaging large households [5], but logistical, geographic, or cultural factors may also play a role. Such misallocation may have an immense negative effect on ITN impact and cost-efficacy as a malaria intervention.

While some studies have observed median net lifespans longer than three years [28, 29, 36, 39], the bulk of existing evidence supports the notion that median net retention is commonly lower than three years [24–27, 30, 38]. The primary motivation for discarding a net in these studies was the perception that it was too torn, with even a modest amount of net damage often regarded as unseemly or untidy [51]. The ITN literature upon which the three-year timeline is founded focuses on nets' anti-mosquito properties, not attrition [52]. This raises a crucial distinction between the technical lifespan of nets based upon their ability to repel and kill mosquitoes despite tearing and other damage, and the effective lifespan of nets, which in many locations seems to be shorter due to lack of physical integrity [Smith et al., in preparation]. Policy solutions to this discrepancy could include more frequent net distributions, shifting to the production and distribution of more physically durable nets [53–58], or using community engagement to change cultural notions about the acceptability of owning and using a torn net. The optimal combination of these strategies will depend heavily on local context and culture.

This analysis finds little evidence of low use among those who own ITNs, especially in areas of high malaria transmission, implying that in most settings there are strong social and cultural norms around the importance of net use when possible. This is consistent with a large body of literature on ITN use rates [13–16]. In this literature, lower use rates are commonly reported in times and geographies of lower actual and perceived risk, such as during the dry season or in highland areas. Use rates in these studies are also highest among children and women, and lowest among teenagers. This analysis supports these seasonal and geographic trends, but does not disaggregate by sex or age. Data from a recent study of net durability across five years and seven countries support the finding that use of available nets remains steady even as net retention declines [59], though the modeling accompanying this analysis suggests a large benefit to increasing the use rate.

Both modeled estimates and national survey data show surprisingly low use rates in Nigeria in the early 2010s and in Ghana from 2005 to the present. The 2013 Nigeria DHS report suggests that this large use gap may have been due to conducting the survey in a season of low malaria transmission, though an independent study also found low rates of net use in Nigeria in the mid-2010s [13]. The 2019 Ghana MIS report cites three primary reasons for lack of net use: that the net was an "extra" or being saved for later, that it was too hot to sleep under a net, or that other anti-mosquito methods were preferred. In both countries use rates are lowest in coastal and urban areas, suggesting correlations with increased wealth, improved housing, lower perceived risk, and access to malaria treatment and alternative malaria prevention strategies.

The combined mechanistic and geospatial modeling framework presented here was necessary to accurately capture spatiotemporal trends in the data, but also lends itself well to the exploration of different distribution scenarios. This feature proved valuable for policy planning in the face of the COVID-19 pandemic, in which

it was necessary to consider counterfactual scenarios of reduced malaria intervention coverage ([60] and Chapter 4). With this framework, it was possible to mechanistically assign unique scenarios to each country to understand how different campaign strategies would affect overall coverage.

This analysis has several limitations. As a continent-level analysis conducted by and for stakeholders based in Europe and North America, this work perpetuates both the "foreign pose" and the "foreign gaze" of global health [61] and should not override the expertise of local researchers and policymakers. Causation is complicated in the relationship between net coverage and malaria– while increased net access and use is known to reduce malaria burden, net distributions are also often targeted to areas where burden is highest. This analysis estimates ITN coverage without incorporating information about malaria transmission in order to allow these estimates to be used as predictors for burden estimation [37, 60]. The coverage metrics reported here do not capture the waning of net efficacy over time due to insecticide decay and physical deterioration, meaning that estimates of ITN use are not fully capturing ITN effectiveness. In the stock and flow model, it is assumed that all nets reported as distributed by NMCPs are successfully provided to households. If, in reality, the net distribution pipeline is leaky and a fraction of nets are not successfully provided, net crop will be overestimated in some years, leading to an underestimate of net retention times. This analysis does not capture differences in net brand or private sector acquisition and distribution of nets [27].

Large data gaps in both space and time impact these results in a number of ways. Conducting the nationally-representative surveys upon which this analysis relies is a Herculean effort, requiring many thousands of work-hours from thousands of people. The cost of such an effort prohibits many countries from conducting these surveys more than once every few years. As such, each year of this analysis includes geospatial information for only a handful of countries and each country

has geospatial information for a median of four years. In addition to increasing the overall uncertainty of this analysis, data sparsity may bias the results in a number of ways. First, the geospatial model assumes that the outcome variables (access deviation, NPC deviation, and use gap) vary smoothly over space, allowing neighboring countries to inform each other's estimates. While this smoothness assumption has some face validity (Figure 3.4), there is no mechanistic reason why this relationship must hold. Second, several countries only have one or two nationally-representative surveys capturing ITN access, and this analysis is unlikely to accurately estimate the ITN time series or LLIN retention times in these cases (see sensitivity analysis in Section 3.3.4). Third, in countries with multiple surveys, data collection commonly occurs in the same season every survey. Since net use is often seasonal, estimates of use rate may be consistently biased towards the use rate of a single season in these countries. Fourth, in this model framework surveys conducted just before large net distributions would lead to predictions of incorrectly low access and use rates, as may have occurred in western Tanzania from 2016 onward. Conducting additional surveys in future years, especially in countries with sparse data at present, would dramatically improve estimates. While net distribution data are far more complete than survey data, they are only provided at the annual and national scale for most countries. Since mass distribution campaigns are typically conducted over several months on a region-by-region schedule, this analysis could gain much precision by incorporating finer-scale space and time data around distributions.

When transforming these estimates into policy decisions, a local understanding of at-risk populations is crucial for effective net distribution. This continental-level analysis defines population-at-risk simplistically as the total population in any pixel with the potential to sustain malaria transmission, when in fact the relative risk of any individual varies based on urbanicity, wealth, housing type, use of other interventions, and local transmission pathways, to name just a few. Additionally, population estimates themselves are often uncertain and may impact results. For

example, the 2020 mass net campaign in Benin found a 13.5% discrepancy between expected and true population [62].

In future iterations of this model framework, improvements will include to incorporating waning net efficacy, taking a more nuanced view of populations-at-risk, and incorporating data (where available) on subnational and sub-annual net distributions. As policy recommendations shift from a focus on universal coverage with ITNs to universal coverage with an appropriate mix of interventions [11], this infrastructure can be used in combination with other mechanistic modeling to advise on optimal intervention packages at national and local scales.

In combination with locally-collected data and the strategic expertise of campaign organizers, these estimates may be able to provide useful context for country programs planning ITN distributions. The evidence provided here of short net retention times implies a potential immense benefit to procuring more durable nets as they reach the market in coming years [53–56, 58], and may suggest the utility of informational campaigns encouraging people to retain slightly torn nets. Countries that are successfully distributing large numbers of nets but experiencing limited gains in access might consider strategy changes to ensure that nets are reaching their intended users. Researchers who may want to adapt this framework to their context can find all code and documentation on GitHub (`https://github.com/bertozzivill/map-itn-cube/tree/publication-2021`), and collaboration on such projects is welcomed.

In malaria policy, understanding the past is a crucial component of planning for the future. This chapter presents maps of ITN access, use, and nets-per-capita from 2000 to 2020 at 5km-by-5km spatial resolution. It shows that ITN distribution campaigns have contributed enormously to coverage in the last two decades, and that the large majority of people who own nets use them. It also shows, however, that current net distribution volume is insufficient to maintain high coverage in a

landscape of low retention rates and inefficient distribution modalities. This flexible and robust modeling framework for net coverage estimation is appropriate for both long-term planning and emergency strategizing in times of crisis, and the author hopes that this tool can continue to play a valuable role in informing ITN policy in years to come.

This chapter provides a detailed description of an ITN coverage modeling framework and the results it estimates from 2000 to 2020. Chapter 4 describes an emergency use case of this framework to estimate how COVID-related service disruptions might impact both ITN coverage and malaria burden in Africa in 2020.

3.7 Additional Plots and Tables

3.7.1 Stock and Flow Model Fit by Country

The plots below show posterior densities and parameter traces from JAGS for the LLIN duration parameter in the stock and flow model (defined as τ in Table 3.4). Countries with wide densities or trace plots that cluster toward the top or bottom of the plot suggest unstable model fits, usually due to sparse survey data.

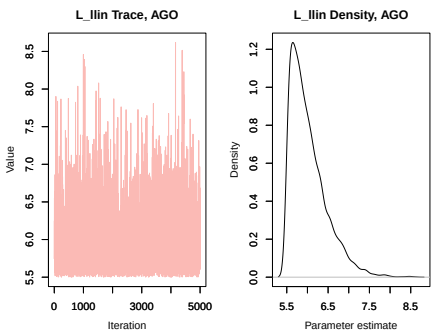


Figure 3.23: Posterior density and parameter trace plots for Angola.

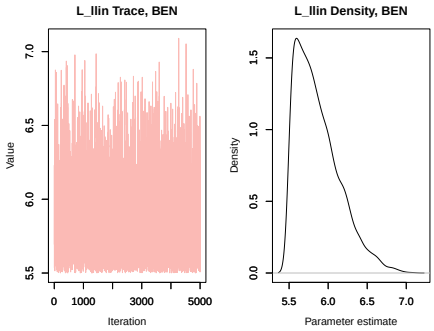


Figure 3.24: Posterior density and parameter trace plots for Benin.

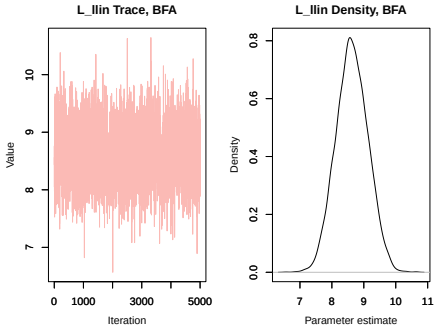


Figure 3.25: Posterior density and parameter trace plots for Burkina Faso.

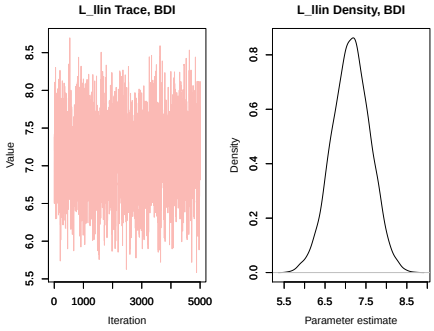


Figure 3.26: Posterior density and parameter trace plots for Burundi.

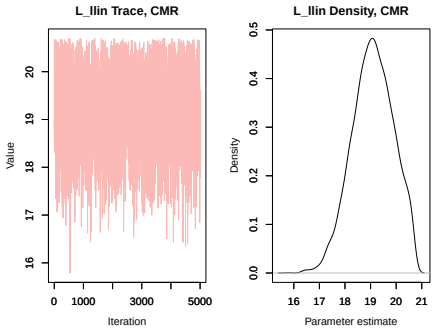


Figure 3.27: Posterior density and parameter trace plots for Cameroon.

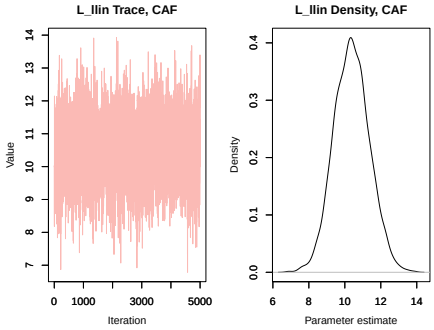


Figure 3.28: Posterior density and parameter trace plots for Central African Republic.

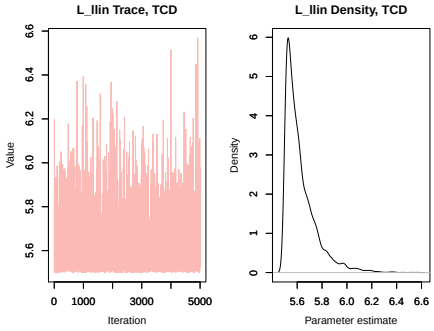


Figure 3.29: Posterior density and parameter trace plots for Chad.

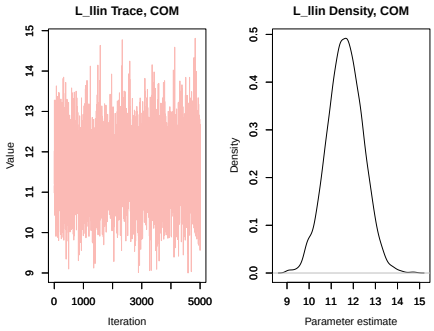


Figure 3.30: Posterior density and parameter trace plots for Comoros.

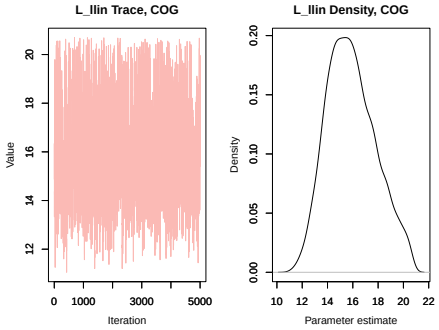


Figure 3.31: Posterior density and parameter trace plots for Congo (Republic of).

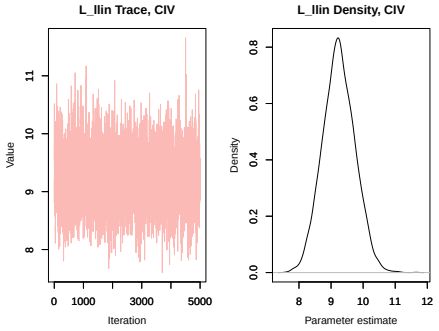


Figure 3.32: Posterior density and parameter trace plots for Cote d'Ivoire.

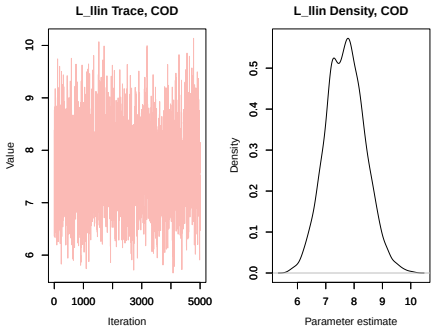


Figure 3.33: Posterior density and parameter trace plots for Democratic Republic of the Congo.

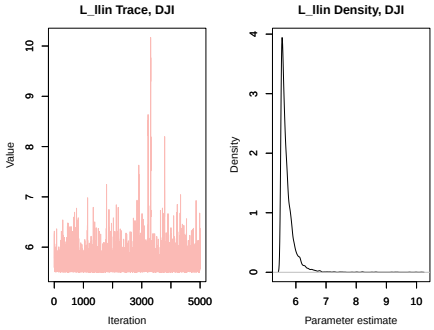


Figure 3.34: Posterior density and parameter trace plots for Djibouti.

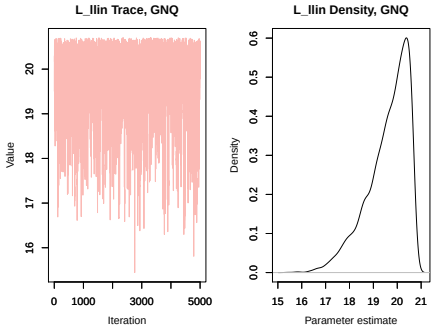


Figure 3.35: Posterior density and parameter trace plots for Equatorial Guinea.

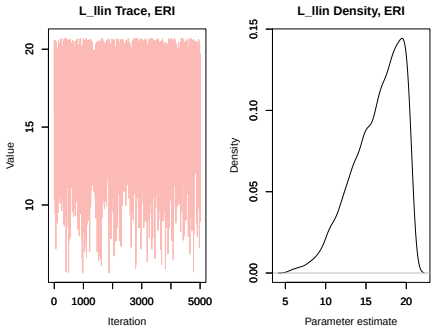


Figure 3.36: Posterior density and parameter trace plots for Eritrea.

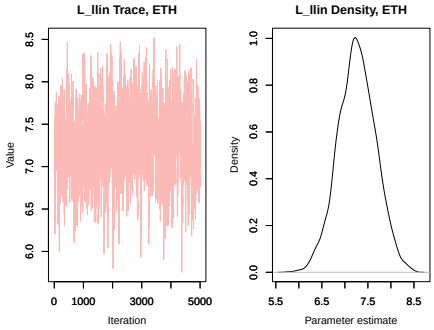


Figure 3.37: Posterior density and parameter trace plots for Ethiopia.

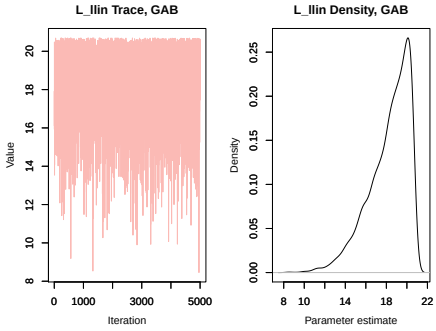


Figure 3.38: Posterior density and parameter trace plots for Gabon.

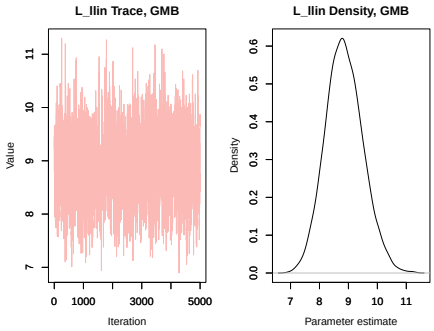


Figure 3.39: Posterior density and parameter trace plots for Gambia.

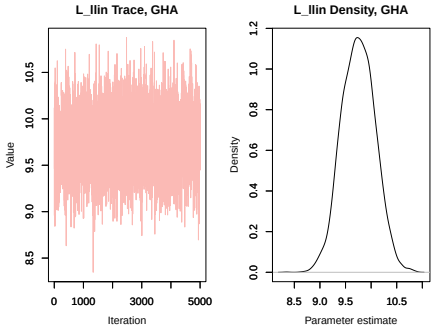


Figure 3.40: Posterior density and parameter trace plots for Ghana.

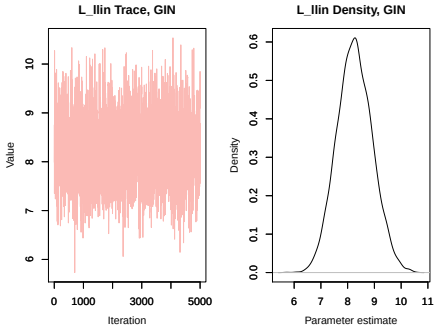


Figure 3.41: Posterior density and parameter trace plots for Guinea.

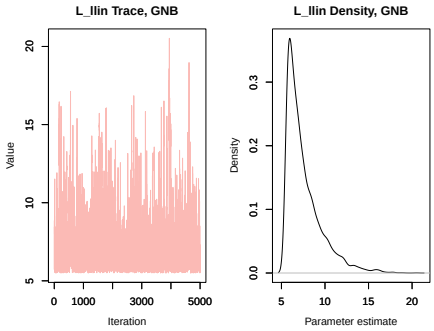


Figure 3.42: Posterior density and parameter trace plots for Guinea-Bissau.

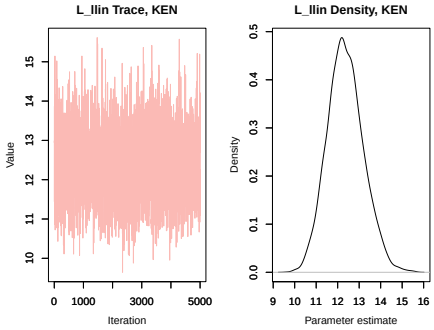


Figure 3.43: Posterior density and parameter trace plots for Kenya.

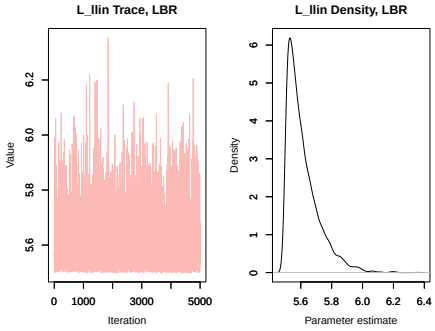


Figure 3.44: Posterior density and parameter trace plots for Liberia.

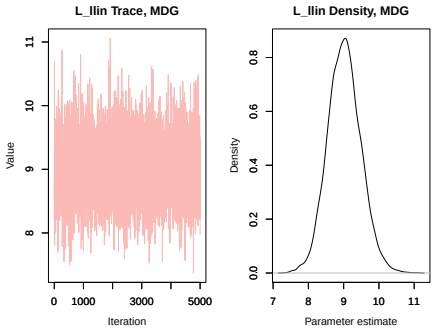


Figure 3.45: Posterior density and parameter trace plots for Madagascar.

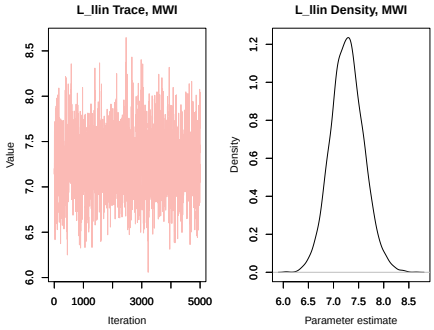


Figure 3.46: Posterior density and parameter trace plots for Malawi.

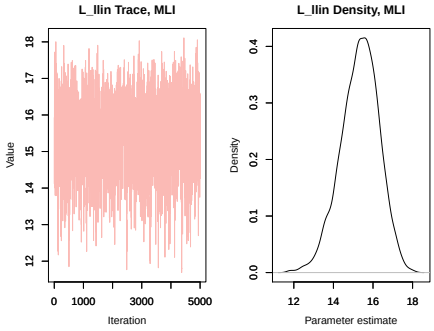


Figure 3.47: Posterior density and parameter trace plots for Mali.

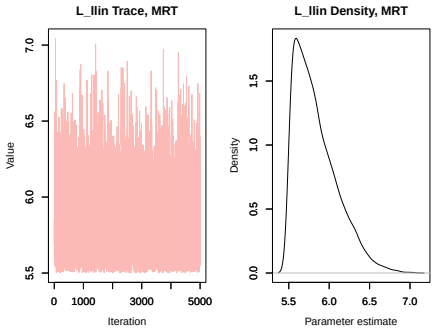


Figure 3.48: Posterior density and parameter trace plots for Mauritania.

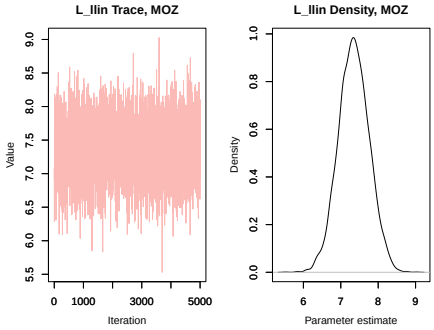


Figure 3.49: Posterior density and parameter trace plots for Mozambique.

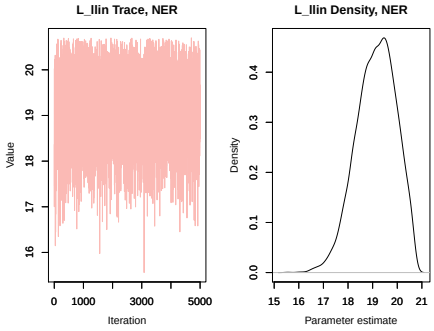


Figure 3.50: Posterior density and parameter trace plots for Niger.

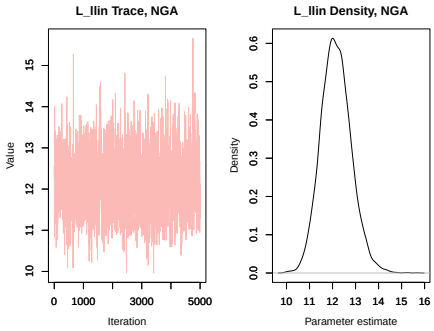


Figure 3.51: Posterior density and parameter trace plots for Nigeria.

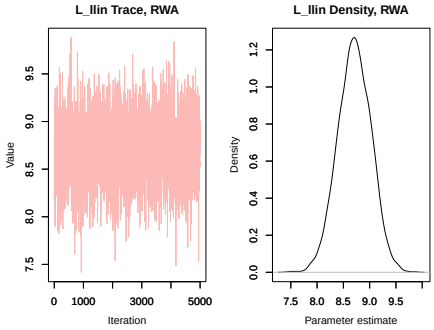


Figure 3.52: Posterior density and parameter trace plots for Rwanda.

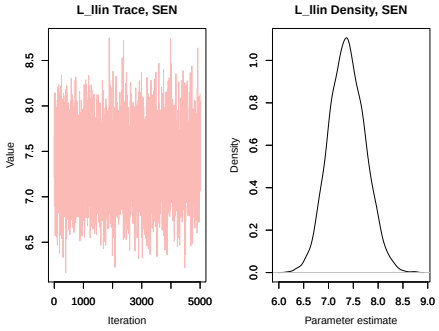


Figure 3.53: Posterior density and parameter trace plots for Senegal.

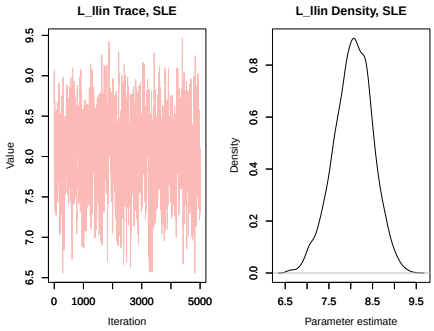


Figure 3.54: Posterior density and parameter trace plots for Sierra Leone.

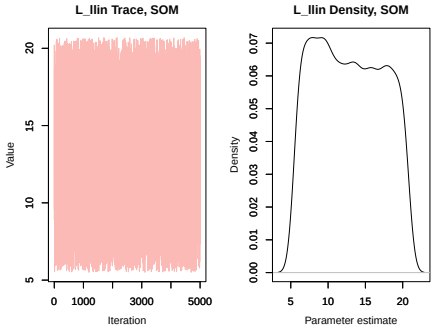


Figure 3.55: Posterior density and parameter trace plots for Somalia.

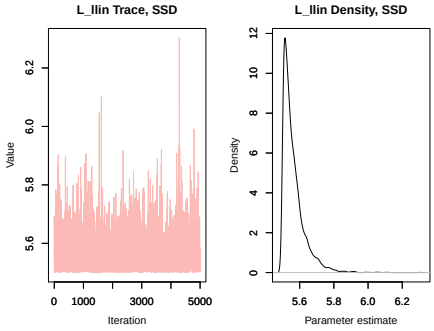


Figure 3.56: Posterior density and parameter trace plots for South Sudan.

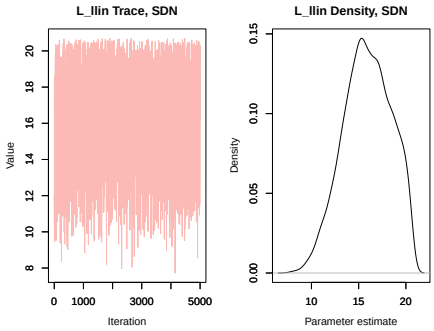


Figure 3.57: Posterior density and parameter trace plots for Sudan.

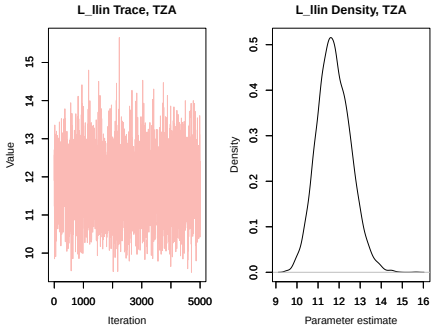


Figure 3.58: Posterior density and parameter trace plots for Tanzania.

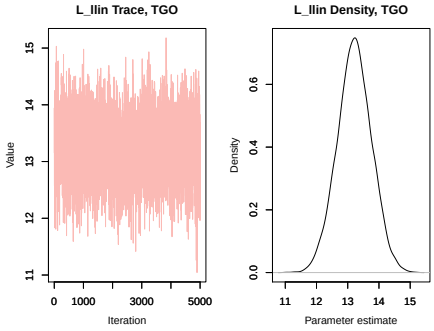


Figure 3.59: Posterior density and parameter trace plots for Togo.

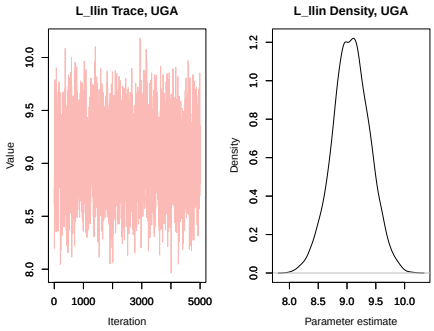


Figure 3.60: Posterior density and parameter trace plots for Uganda.

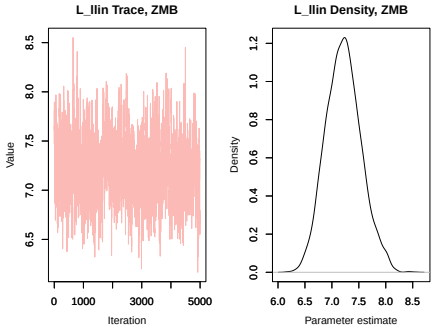


Figure 3.61: Posterior density and parameter trace plots for Zambia.

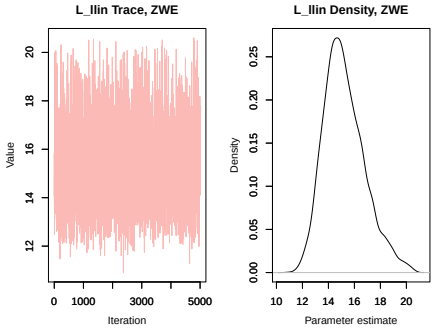


Figure 3.62: Posterior density and parameter trace plots for Zimbabwe.

3.7.2 Stock and Flow Sensitivity Analysis by Country

Plotted below are model fits by country for the stock and flow sensitivity analysis. Columns indicate the type of sensitivity analysis, and rows indicate how many surveys were included in model fitting. Blue curves represent LLIN model estimates, while red curves represent cITN model estimates. Solid curves show mean values, and shaded areas represent 95% credible intervals. Mean values for survey data are shown in circles for LLINs and squares for cITNs, with vertical lines showing the standard error of the survey. Surveys used for model fitting are shown in color, while out-of-sample surveys are shown in black. The RMSEs printed on each plot are for LLINs only.

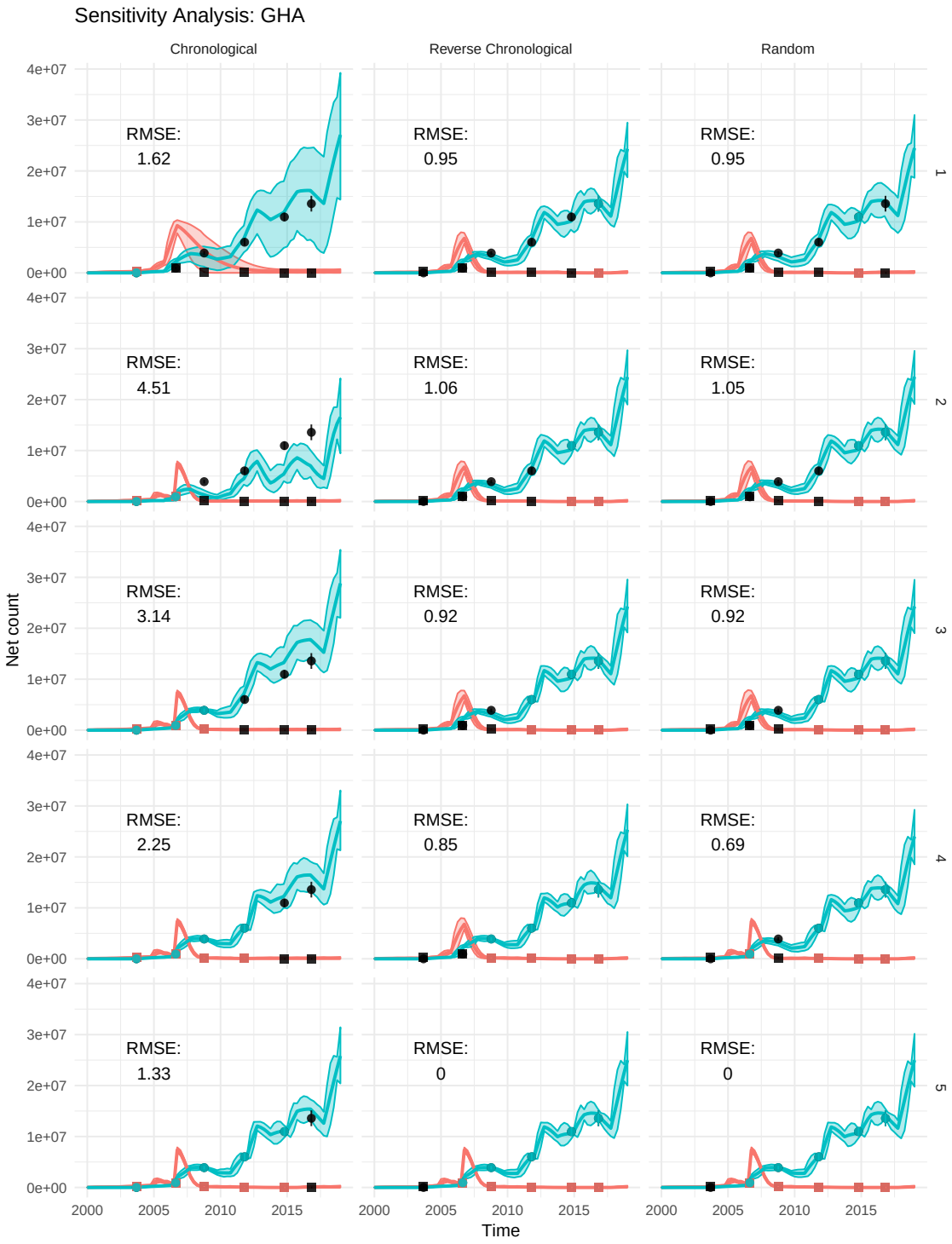


Figure 3.63: Sensitivity analysis results for Ghana.

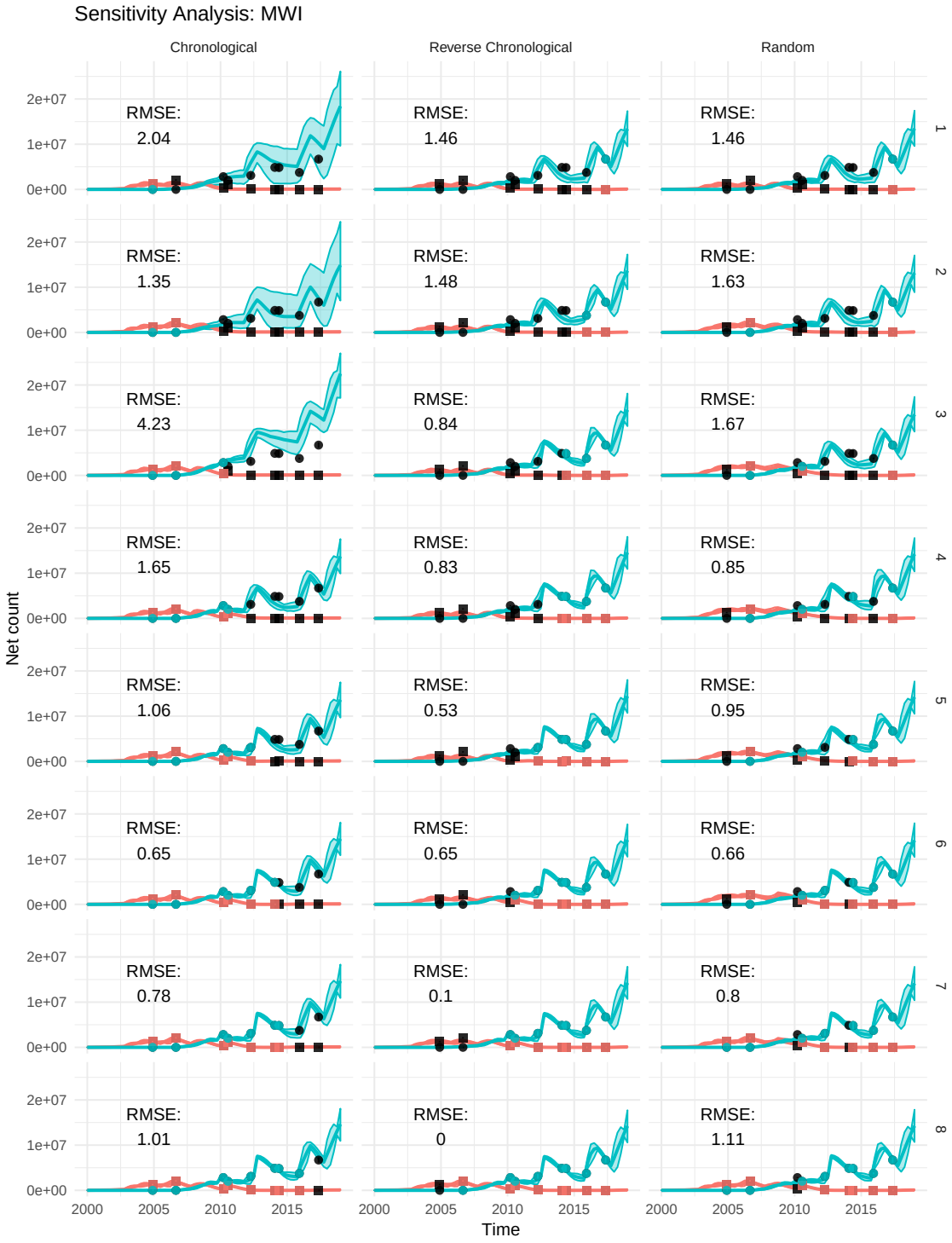


Figure 3.64: Sensitivity analysis results for Malawi.

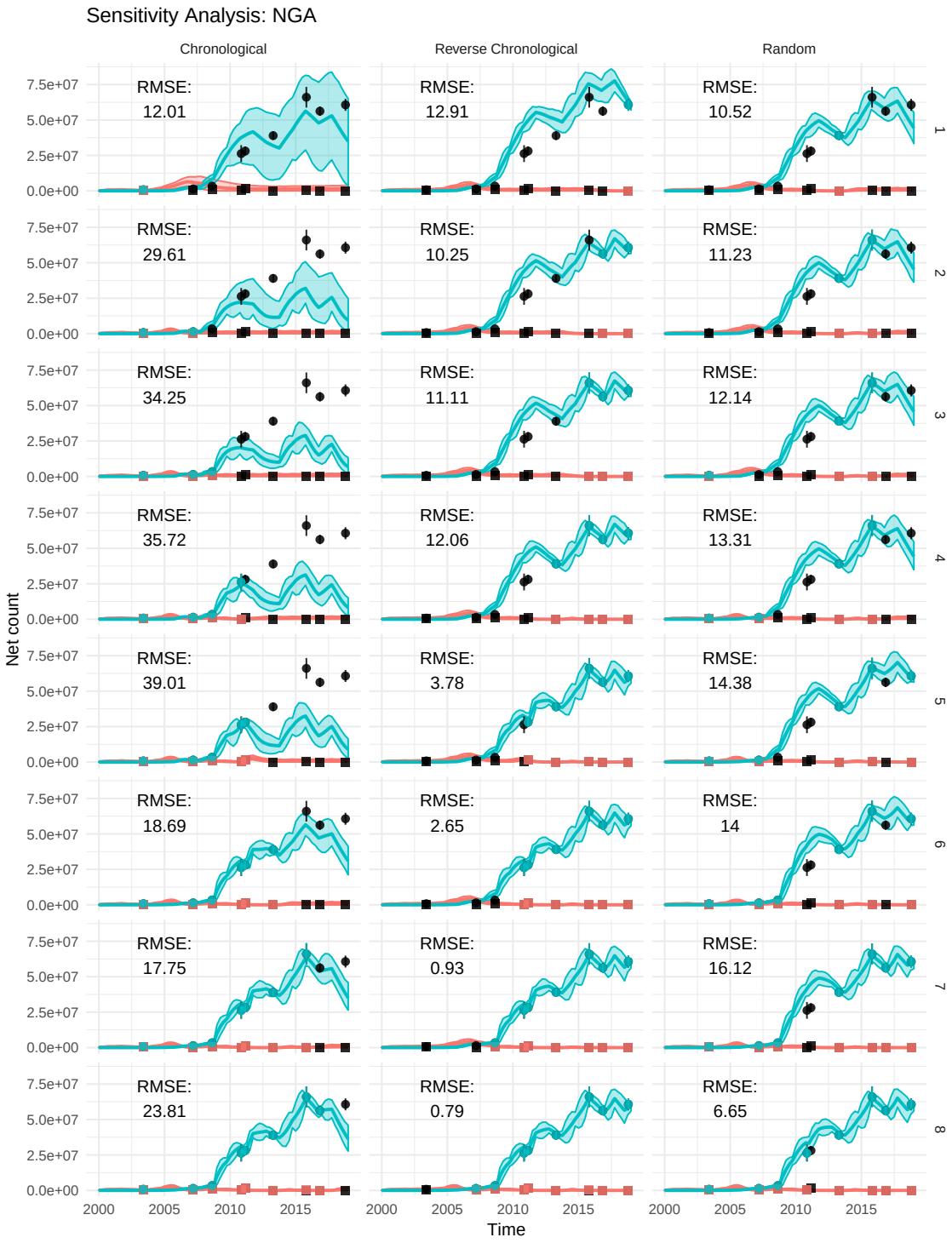


Figure 3.65: Sensitivity analysis results for Nigeria.

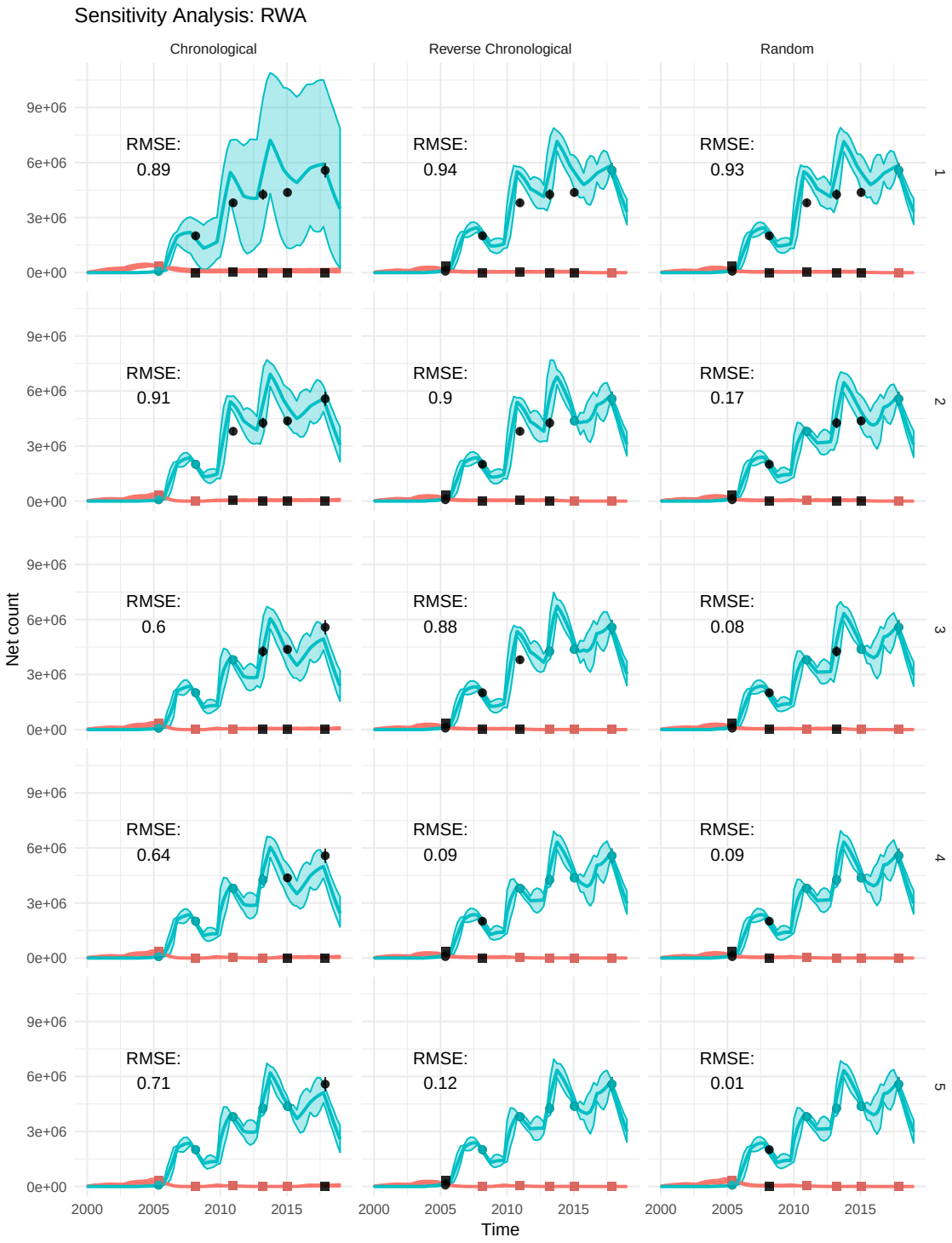


Figure 3.66: Sensitivity analysis results for Rwanda.

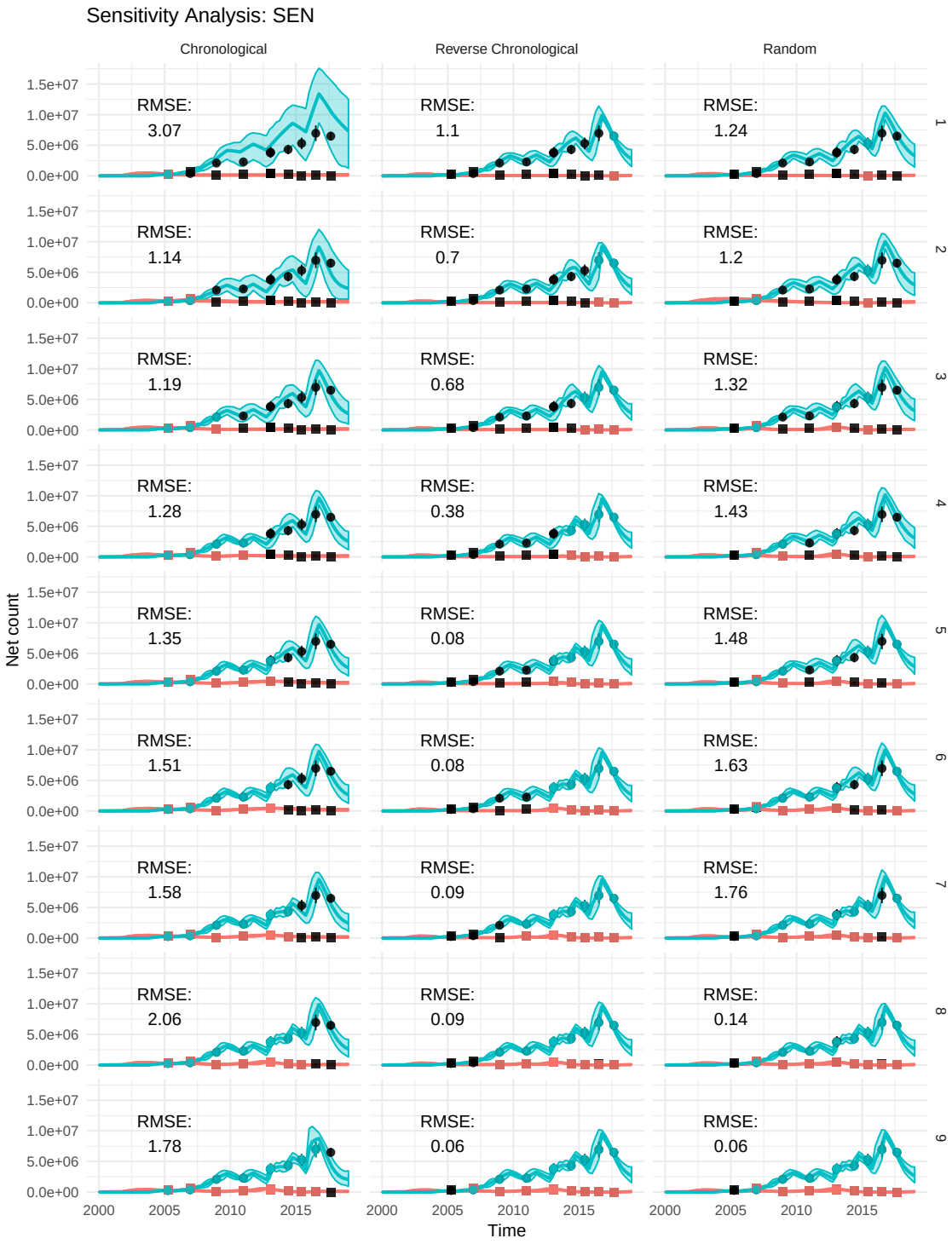


Figure 3.67: Sensitivity analysis results for Senegal.

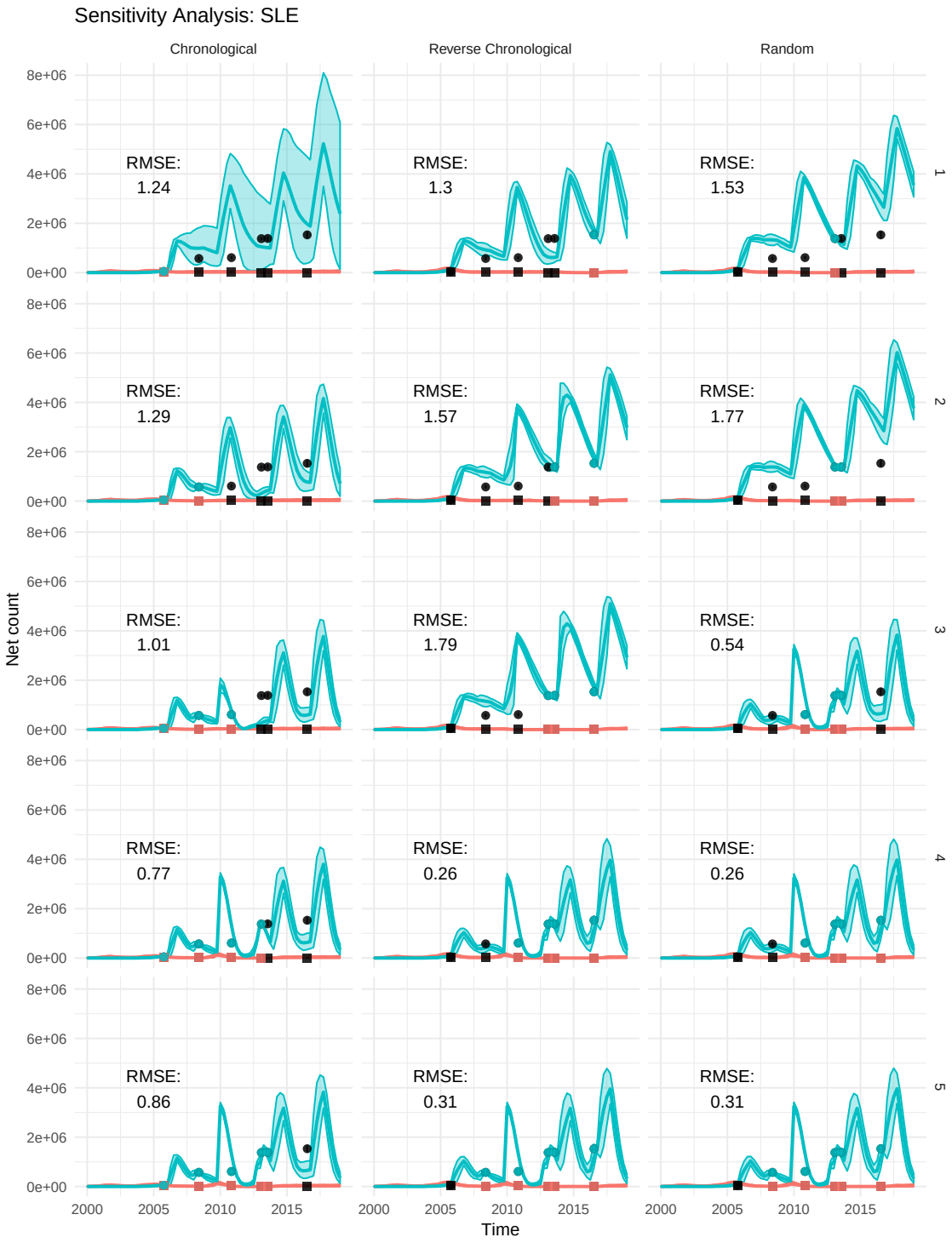


Figure 3.68: Sensitivity analysis results for Sierra Leone.

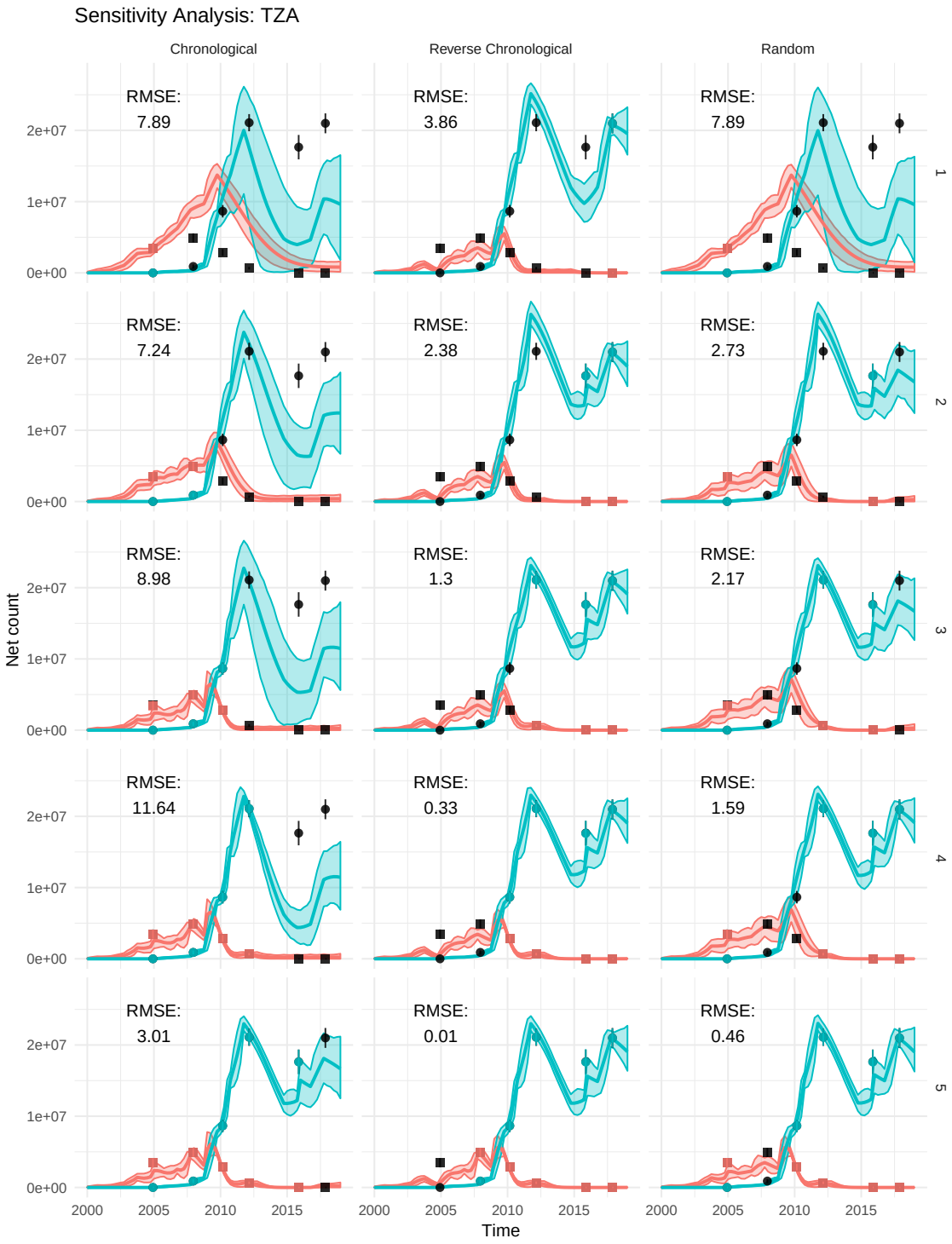


Figure 3.69: Sensitivity analysis results for Tanzania.

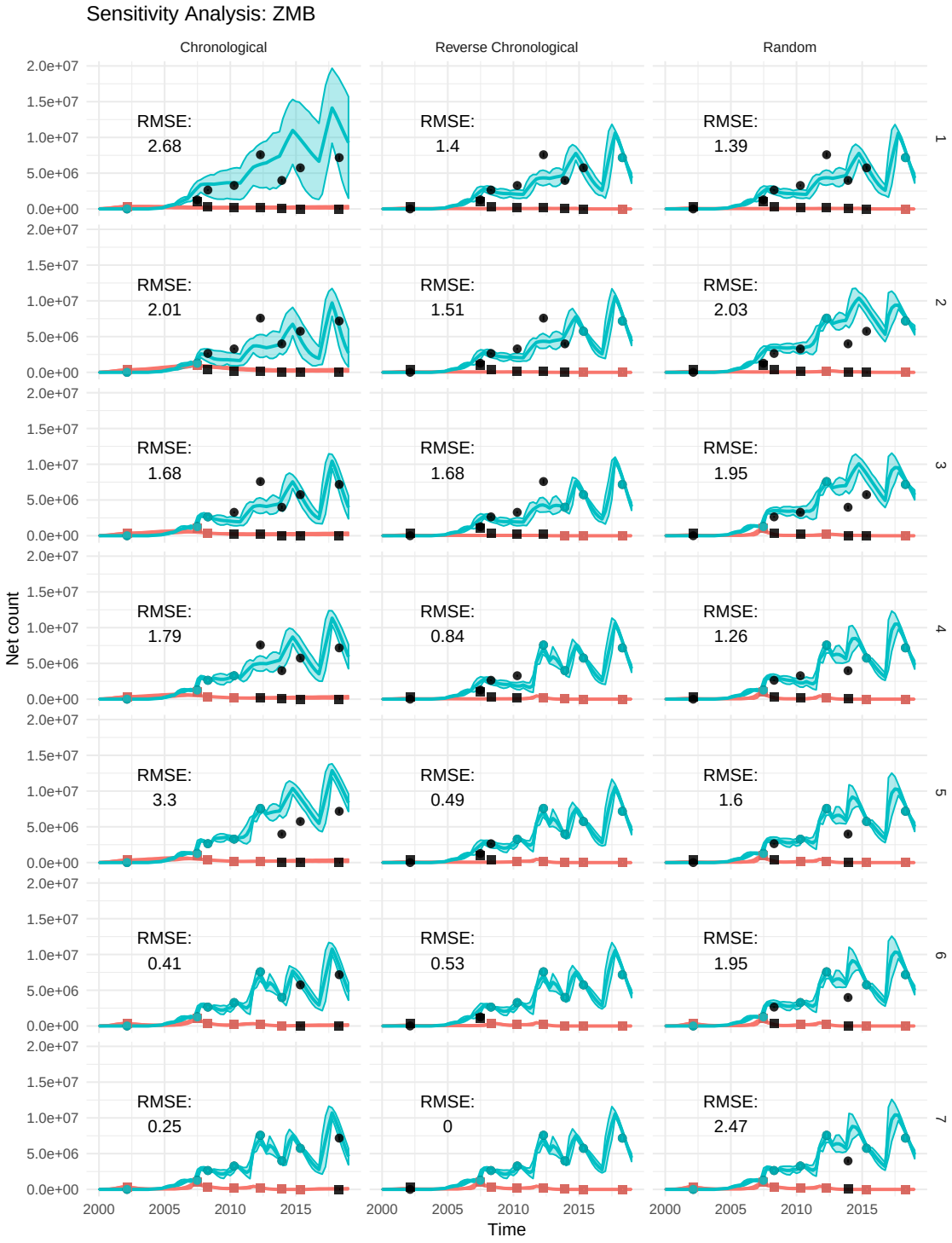


Figure 3.70: Sensitivity analysis results for Zambia.

3.7.3 Sources for Report-Only Surveys

Some surveys were not available at the household level, so summary measures were extracted from survey reports. This section details those surveys and the locations and page numbers in which the data were extracted.

Report-only survey details.

Country	ISO3	Survey Years	Source	Report Page	Link
Burkina Faso	BFA	2006	MICS3	6	https://mics-surveys-prod.s3.amazonaws.com/MICS3/West%20and%20Central%20Africa/Burkina%20Faso/2006/Final/Burkina%20Faso%202006%20MICS_French.pdf
Burundi	BDI	2005	MICS3	24	https://mics-surveys-prod.s3.amazonaws.com/MICS3/Eastern%20and%20Southern%20Africa/Burundi/2005/Final/Burundi%202005%20MICS_French.pdf
Central African Republic	CAR	2006	MICS3	30	https://mics-surveys-prod.s3.amazonaws.com/MICS3/West%20and%20Central%20Africa/Central%20African%20Republic/2006/Final/Central%20African%20Republic%202006%20MICS_French.pdf
Code d'Ivoire	CIV	2006	MICS3	26	https://mics-surveys-prod.s3.amazonaws.com/MICS3/West%20and%20Central%20Africa/C%C3%B4te%20d%27Ivoire/2006/Final/Cote%20d%27Ivoire%202006%20MICS_French.pdf

Continued on next page

Report-only survey details, cont.					
Country	**ISO3**	**Survey Years**	**Source**	**Report Page**	**Link**
Djibouti	DJI	2006	MICS3	25	https://mics-surveys-prod.s3.amazonaws.com/MICS3/Middle%20East%20and%20North%20Africa/Djibouti/2006/Final/Djibouti%202006%20MICS_French.pdf
Djibouti	DJI	2009	MIS	15	https://malariaatlas.org/media/PDF/Djibouti_MIS_Report.pdf
Ethiopia	ETH	2007	MIS	10	https://www.cartercenter.org/resources/pdfs/news/health_publications/malaria/Ethiopia_2007_MIS_Report.pdf
Ethiopia	ETH	2011	MIS	12	https://www.ephi.gov.et/images/downloads/ethiopia-mis2011-report.pdf
Ethiopia	ETH	2015	MIS	20	https://www.malariasurveys.org/documents/Ethiopia_MIS_2015.pdf
Equatorial Guinea	GNQ	2011	DHS	17	https://dhsprogram.com/pubs/pdf/FR271/FR271.pdf
Gambia	GMB	2005-6	MICS3	15	https://mics-surveys-prod.s3.amazonaws.com/MICS3/West%20and%20Central%20Africa/Gambia/2005-2006/Final/Gambia%202005-06%20MICS_English.pdf
Gambia	GMB	2014	MIS	18	http://ccmgambia.org/images/pdfs/Final-Report-MIS-2014.pdf
Gambia	GMB	2017	MIS	13	https://www.malariasurveys.org/documents/GMIS%202017%20final%20report.pdf

Continued on next page

Report-only survey details, cont.

Country	ISO3	Survey Years	Source	Report Page	Link
Ghana	GHA	2006	MICS3	8	https://mics-surveys-prod.s3.amazonaws.com/MICS3/West%20and%20Central%20Africa/Ghana/2006/Final/Ghana%202006%20MICS_English.pdf
Guinea-Bissau	GNB	2006	MICS3	30	https://mics-surveys-prod.s3.amazonaws.com/MICS3/West%20and%20Central%20Africa/Guinea-Bissau/2006/Final/Guinea-Bissau%202006%20MICS_French.pdf
Malawi	MWI	2006	MICS3	10	https://mics-surveys-prod.s3.amazonaws.com/MICS3/Eastern%20and%20Southern%20Africa/Malawi/2006/Final/Malawi%202006%20MICS_English.pdf
Malawi	MWI	2010	MIS	13	https://files.givewell.org/files/DWDA%202009/AMF/Malawi_MIS_2010_Final.pdf
Mauritania	MRT	2007	MICS3	10	http://mics.unicef.org/surveys
Nigeria	NGA	2007	MICS3	6	https://mics-surveys-prod.s3.amazonaws.com/MICS3/West%20and%20Central%20Africa/Nigeria/2007/Final/Nigeria%202007%20MICS_English.pdf
Senegal	SEN	2013	DHS	12	https://dhsprogram.com/pubs/pdf/FR288/FR288eng.pdf
Sierra Leone	SLE	2005	MICS3	7	https://mics.unicef.org/surveys
Sierra Leone	SLE	2013	MIS	18	https://dhsprogram.com/pubs/pdf/MIS15/MIS15.pdf

Continued on next page

Report-only survey details, cont.					
Country	**ISO3**	**Survey Years**	**Source**	**Report Page**	**Link**
Somalia	SOM	2006	MICS3	23	https://mics-surveys-prod.s3.amazonaws.com/MICS3/Eastern%20and%20Southern%20Africa/Somalia/2006/Final/Somalia%202006%20MICS_English.pdf
South Sudan	SSD	2009	MIS	9	https://www.malariasurveys.org/documents/SS%202009_Malaria%20Indicator%20Survey%20Report.pdf
South Sudan	SSD	2013	MIS	8	https://www.malariasurveys.org/documents/MIS%20Report%202013%20Final%20Pages.pdf
South Sudan	SSD	2017	MIS		https://www.malariasurveys.org/documents/SSMIS%202017%20final%20report%202019%2005%2020_signed%20(1).pdf
Togo	TGO	2006	MICS3	7	https://mics-surveys-prod.s3.amazonaws.com/MICS3/West%20and%20Central%20Africa/Togo/2006/Final/Togo%202006%20MICS_French.pdf
Zambia	ZMB	2008	MIS	8	https://www.malariasurveys.org/documents/Zambia_MIS_2008.pdf
Zambia	ZMB	2010	MIS	9	https://www.malariasurveys.org/documents/Zambia_MIS_2010.pdf
Zambia	ZMB	2012	MIS	8	https://www.malariasurveys.org/documents/Zambia_MIS_2012.pdf

Continued on next page

Report-only survey details, cont.					
Country	**ISO3**	**Survey Years**	**Source**	**Report Page**	**Link**
Zambia	ZMB	2015	MIS	8	https://static1.squarespace.com/static/58d002f017bffcf99fe21889/t/5e348f9c95862d328a90cf42/1580502944434/Zambia-MIS2015_Jan20-nosigs.pdf
Zambia	ZMB	2018	MIS	18	https://www.malariasurveys.org/documents/Zambia_MIS2018_FINAL-withSigs_20190415.pdf
Zimbabwe	ZWE	2009	MICS3	10	https://mics-surveys-prod.s3.amazonaws.com/MICS3/Eastern%20and%20Southern%20Africa/Zimbabwe/2009/Final/Zimbabwe%202009%20MICS_English.pdf

Table 3.13: Sources for those surveys used in ITN models in which only report-level data were available.

Bibliography

1. Bhatt, S. *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* **526,** 207–211. ISSN: 14764687 (2015).
2. Yukich, J. O. *et al.* Costs and consequences of large-scale vector control for malaria. *Malaria Journal* **7,** 258. ISSN: 14752875. http://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-7-258 (2008).
3. Wisniewski, J., Acosta, A., Kolaczinski, J., Koenker, H. & Yukich, J. Systematic review and meta-analysis of the cost and cost-effectiveness of distributing insecticide-treated nets for the prevention of malaria. *Acta Tropica* **202,** 105229. ISSN: 18736254. https://doi.org/10.1016/j.actatropica.2019.105229 (2020).
4. Kilian, A., Koenker, H. & Paintain, L. Estimating population access to insecticide-treated nets from administrative data: Correction factor is needed. *Malaria Journal* **12,** 1–10. ISSN: 14752875 (2013).
5. Koenker, H. *et al.* Assessing whether universal coverage with insecticide-treated nets has been achieved: Is the right indicator being used? *Malaria Journal* **17,** 355. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/s12936-018-2505-0 (2018).
6. Global Malaria Programme. *Achieving and maintaining universal coverage with long-lasting insecticidal nets for malaria control* tech. rep. (2017).
7. Gething, P. W. *et al.* A new world malaria map: Plasmodium falciparum endemicity in 2010. *Malaria Journal* **10,** 378. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-10-378 (2011).
8. WHO. *World Malaria Report 2008* tech. rep. (World Health Organization, 2008).
9. Global Malaria Programme. *INSECTICIDE-TREATED MOSQUITO NETS: a WHO Position Statement* tech. rep. (2007). www.medlive.cn.
10. WHO. *World Malaria Report 2010* tech. rep. (2010).
11. Global Malaria Programme & WHO. *WHO Malaria Policy Advisory Committee (MPAC) meeting: Meeting Report* tech. rep. October (2019), 1–23. http://www.who.int/malaria/mpac/mpacmembers/en/.

12. Bhatt, S. *et al.* Coverage and system efficiencies of insecticide-treated nets in Africa from 2000 to 2017. *eLife* **4.** ISSN: 2050084X (2015).

13. Storey, J. D. *et al.* Associations between ideational variables and bed net use in Madagascar, Mali, and Nigeria. *BMC Public Health* **18,** 484. ISSN: 14712458. https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-018-5372-2 (2018).

14. VectorWorks. *ITN Access and Use Mapping net culture in sub-Saharan Africa* tech. rep. ().

15. Shah, M. P. *et al.* The effectiveness of older insecticide-treated bed nets (ITNs) to prevent malaria infection in an area of moderate pyrethroid resistance: Results from a cohort study in Malawi. *Malaria Journal* **19,** 24. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/s12936-020-3106-2 (2020).

16. Koenker, H. *et al.* Quantifying seasonal variation in insecticide-treated net use among those with access. *American Journal of Tropical Medicine and Hygiene* **101,** 371–382. ISSN: 00029637 (2019).

17. Koenker, H. *et al.* Trends in weekly reported net use by children during and after rainy season in central Tanzania. *Malaria Journal* **11,** 218. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-11-218 (2012).

18. Rek, J. *et al.* Non-adherence to long-lasting insecticide treated bednet use following successful malaria control in Tororo, Uganda. *PLOS ONE* **15** (ed Carvalho, L. H.) e0243303. ISSN: 1932-6203. https://dx.plos.org/10.1371/journal.pone.0243303 (2020).

19. Olapeju, B. *et al.* Age and gender trends in insecticide-treated net use in sub-Saharan Africa: A multi-country analysis 11 Medical and Health Sciences 1117 Public Health and Health Services. *Malaria Journal* **17,** 423. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/s12936-018-2575-z (2018).

20. Short, R., Gurung, R., Rowcliffe, M., Hill, N. & Milner-Gulland, E. J. The use of mosquito nets in fisheries: A global perspective. *PLOS ONE* **13** (ed Munderloh, U. G.) e0191519. ISSN: 1932-6203. https://dx.plos.org/10.1371/journal.pone.0191519 (2018).

21. PMI VectorWorks. *Identifying and Mitigating Misuse of Insecticide-Treated Nets for Fishing Toolkit* tech. rep. (2018).

22. Berthe, S. *et al.* Poverty and food security: Drivers of insecticide-treated mosquito net misuse in Malawi. *Malaria Journal* **18,** 320. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/s12936-019-2952-2 (2019).

23. Santos, E. M. *et al.* "After those nets are torn, most people use them for other purposes": An examination of alternative bed net use in western Kenya. *Malaria Journal* **19,** 272. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/s12936-020-03342-1 (2020).

24. Gnanguenon, V., Azondekon, R., Oke-Agbo, F., Beach, R. & Akogbeto, M. Durability assessment results suggest a serviceable life of two, rather than three, years for the current long-lasting insecticidal (mosquito) net (LLIN) intervention in Benin. *BMC Infectious Diseases* **14,** 69. ISSN: 14712334. https://bmcinfectdis.biomedcentral.com/articles/10.1186/1471-2334-14-69 (2014).

25. Solomon, T. *et al.* Bed nets used to protect against malaria do not last long in a semi-arid area of Ethiopia: A cohort study. *Malaria Journal* **17,** 239. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/s12936-018-2391-5 (2018).

26. Lorenz, L. M. *et al.* Comparative functional survival and equivalent annual cost of 3 long-lasting insecticidal net (LLIN) products in Tanzania: A randomised trial with 3-year follow up. *PLOS Medicine* **17** (ed Ashley, E. A.) e1003248. ISSN: 1549-1676. https://dx.plos.org/10.1371/journal.pmed.1003248 (2020).

27. Mansiangi, P. *et al.* Comparing the durability of the long-lasting insecticidal nets DawaPlus® 2.0 and DuraNet© in northwest Democratic Republic of Congo. *Malaria Journal* **19,** 189. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/s12936-020-03262-0 (2020).

28. Obi, E. *et al.* Monitoring the physical and insecticidal durability of the long-lasting insecticidal net DawaPlus® 2.0 in three States in Nigeria. *Malaria Journal* **19,** 124. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/s12936-020-03194-9 (2020).

29. Kilian, A. *et al.* Field durability of the same type of long-lasting insecticidal net varies between regions in Nigeria due to differences in household behaviour and living conditions. *Malaria Journal* **14,** 123. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/s12936-015-0640-4 (2015).

30. PMI. *Durability Monitoring of LLINs in Mozambique* tech. rep. (PMI, 2018). https://www.pmi.gov/docs/default-source/default-document-library/pmi-reports/durability-monitoring-of-llin-in-mozambique-final-report-after-36-months-follow-up-2018.pdf?sfvrsn=4.

31. Flaxman, A. D. *et al.* Rapid scaling up of insecticide-treated bed net coverage in Africa and its relationship with development assistance for health: A systematic synthesis of supply, distribution, and household survey data. *PLoS Medicine* **7.** ISSN: 15491277 (2010).

32. Kim, H. *et al.* Spatiotemporal analysis of insecticide-treated net use for children under 5 in relation to socioeconomic gradients in Central and East Africa. *Malaria Journal* **19,** 163. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/s12936-020-03236-2 (2020).

33. Noor, A. M., Mutheu, J. J., Tatem, A. J., Hay, S. I. & Snow, R. W. Insecticide-treated net coverage in Africa: mapping progress in 2000-07. *The Lancet* **373,** 58–67. ISSN: 01406736 (2009).

34. Koenker, H. & Kilian, A. Recalculating the Net Use Gap: A Multi-Country Comparison of ITN Use versus ITN Access. *PLoS ONE* **9** (ed Gatton, M. L.) e97496. ISSN: 1932-6203. https://dx.plos.org/10.1371/journal.pone.0097496 (2014).

35. Yukich, J. *et al.* Planning long lasting insecticide treated net campaigns: Should households' existing nets be taken into account? *Parasites and Vectors* **6,** 174. ISSN: 17563305. https://parasitesandvectors.biomedcentral.com/articles/10.1186/1756-3305-6-174 (2013).

36. Hakizimana, E. *et al.* Monitoring long-lasting insecticidal net (LLIN) durability to validate net serviceable life assumptions, in Rwanda. *Malaria Journal* **13,** 344. ISSN: 14752875. http://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-13-344 (2014).

37. Weiss, D. J. *et al.* Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum, 2000–17: a spatial and temporal modelling study. *The Lancet* **394,** 322–331. ISSN: 1474547X (2019).

38. Tan, K. R. *et al.* A longitudinal study of the durability of long-lasting insecticidal nets in Zambia. *Malaria Journal* **15,** 106. ISSN: 14752875. http://www.malariajournal.com/content/15/1/106 (2016).

39. Kilian, A. *et al.* Evidence for a useful life of more than three years for a polyester-based long-lasting insecticidal mosquito net in Western Uganda. *Malaria Journal* **10,** 299. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-10-299 (2011).

40. Kilian, A., Boulay, M., Koenker, H. & Lynch, M. How many mosquito nets are needed to achieve universal coverage? Recommendations for the quantification and allocation of long-lasting insecticidal nets for mass campaigns. *Malaria Journal* **9,** 330. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-9-330 (2010).

41. Singh, J. A Characterization of Positive Poisson Distribution and its Statistical Application. *SIAM Journal on Applied Mathematics* **34,** 545–548 (1978).

42. Weiss, D. J. *et al.* An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS Journal of Photogrammetry and Remote Sensing* **98,** 106–118. ISSN: 09242716. /pmc/articles/PMC4308023/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4308023/ (2014).

43. Lindgren, F., Rue, H. & Lindström, J. An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **73,** 423–498. ISSN: 13697412. https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1467-9868.2011.00777.x%20https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2011.00777.x%20https://rss.onlinelibrary.wiley.com/doi/10.1111/j.1467-9868.2011.00777.x (2011).

44. Burbidge, J. B., Magee, L. & Robb, A. L. Alternative Transformations to Handle Extreme Values of the Dependent Variable. *Journal of the American Statistical Association* **83,** 123. ISSN: 01621459 (1988).

45. Winch, P. J. *et al.* Seasonal variation in the perceived risk of malaria: Implications for the promotion of insecticide-impregnated bed nets. *Social Science and Medicine* **39,** 63–75. ISSN: 02779536 (1994).

46. Battle, K. E. *et al.* Mapping the global endemicity and clinical burden of Plasmodium vivax, 2000–17: a spatial and temporal modelling study. *The Lancet* **394,** 332–343. ISSN: 1474547X. http://dx.doi.org/10.1016/ (2019).

47. Fernando, S. D. *et al.* Sleeping arrangements under long-lasting impregnated mosquito nets: differences during low and high malaria transmission seasons. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **103,** 1204–1210. ISSN: 00359203 (2009).

48. WHO. *World Malaria Report 2013* tech. rep. (2013). https://www.who.int/malaria/publications/world%7B%5C_%7Dmalaria%7B%5C_%7Dreport%7B%5C_%7D2013/en/.

49. WHO. *World Malaria Report 2020* tech. rep. December (2020), 238.

50. Winskill, P., Walker, P. G., Griffin, J. T. & Ghani, A. C. Modelling the cost-effectiveness of introducing the RTS,S malaria vaccine relative to scaling up other malaria interventions in sub-Saharan Africa. *BMJ Global Health* **2,** 90. https://gh.bmj.com/content/2/1/e000090 (2017).

51. Koenker, H. *et al.* What happens to lost nets: A multi-country analysis of reasons for LLIN attrition using 14 household surveys in four countries. *Malaria Journal* **13,** 464. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-13-464 (2014).

52. WHO. *Guidelines for laboratory and field-testing of long-lasting insecticidal nets* tech. rep. (2013). www.who.int.

53. Okumu, F. *The fabric of life: What if mosquito nets were durable and widely available but insecticide-free?* 2020. https://malariajournal.biomedcentral.com/articles/10.1186/s12936-020-03321-6.

54. Wheldrake, A., Guillemois, E., Arouni, H., Chetty, V. & Russell, S. J. The causes of holes and loss of physical integrity in long-lasting insecticidal nets. *Malaria Journal* **20,** 1–10. ISSN: 14752875. https://doi.org/10.1186/s12936-020-03567-0 (2021).

55. Wheldrake, A., Guillemois, E., Chetty, V., Kilian, A. & Russell, S. J. Development of a single resistance to damage metric for mosquito nets related to physical integrity in the field. *Malaria Journal* **20,** 1–13. ISSN: 14752875. https://doi.org/10.1186/s12936-020-03566-1 (2021).

56. Wheldrake, A., Guillemois, E., Arouni, H., Chetty, V. & Russell, S. J. Textile testing to assess the resistance to damage of long-lasting insecticidal nets for malaria control and prevention. *Malaria Journal* **20,** 1–17. ISSN: 14752875. https://doi.org/10.1186/s12936-020-03571-4 (2021).

57. Kilian, A. *et al.* Correlation of textile 'resistance to damage' scores with actual physical survival of long-lasting insecticidal nets in the field. *Malaria Journal* **20,** 1–10. ISSN: 14752875. https://doi.org/10.1186/s12936-020-03570-5 (2021).

58. Russell, S. J., Wheldrake, A. & Chetty, V. *Modes of physical deterioration of LLINs in different settings* tech. rep. (6th Meeting of the Durability of LLIN in the Field Work Stream of the Roll Back Malaria Vector Control Working Group, 2014).

59. Briet, O. *et al.* Attrition, physical integrity and insecticidal activity of long-lasting insecticidal nets in sub-Saharan Africa and modelling of their impact on vectorial capacity. *Malaria journal* **19,** 310. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/s12936-020-03383-6 (2020).

60. Weiss, D. J. *et al.* Indirect effects of the COVID-19 pandemic on malaria intervention coverage, morbidity, and mortality in Africa: a geospatial modelling analysis. *The Lancet Infectious Diseases* (2020).

61. Abimbola, S. *The foreign gaze: Authorship in academic global health* 2019. http://gh.bmj.com/.

62. Aïkpon, R. *et al.* Digitalized mass distribution campaign of insecticide-treated nets (ITNs) in the particular context of Covid-19 pandemic in Benin: challenges and lessons learned. *Malaria Journal* **19,** 431. ISSN: 14752875. https://malariajournal.biomedcentral.com/articles/10.1186/s12936-020-03508-x (2020).

Chapter 4

Predicting the Impact of COVID-Related Service Disruptions on Insecticide-Treated Net Coverage, 2020

> I want it to mean that the garden in which we work when we are no longer optimistic is not a retreat from the world, but a place where we cultivate the world.
>
> Eula Biss
> *On Immunity*

4.1 Introduction

In early 2020, the COVID-19 pandemic shook the globe. In a landscape of uncertainty and panic, no previously-constructed plan for the year remained intact, and malaria service provision was no exception. In fact, the malaria situation was particularly fragile for a number of reasons both structural and coincidental. Because malaria is closely tied to environmental and entomological conditions, in most malaria-endemic areas transmission is dampened only through active interventions

such as vector control and antimalarial drugs. In the absence of these barriers, transmission can and will resurge [1]. Thus, as a disease it is sensitive to service infrastructure shocks at any time [2]. In a deeply unfortunate coincidence, the COVID pandemic took hold in the same year that a record number of mass Insecticide-Treated Net (ITN) distribution campaigns were scheduled to take place in sub-Saharan Africa [3]. In the early days of the pandemic, many countries considered canceling or delaying their campaigns to help minimize COVID transmission– a prudent and thoughtful consideration that nonetheless put access to a crucial malaria prevention tool at risk for millions of people. At the same time that net distribution campaigns were being weighed in the balance, access to life-saving malaria medications was also under threat. Antimalarial drug prices surged in the spring of 2020 as some speculated that these drugs might be effective treatments for COVID [4]. Simultaneously, some drug companies were reducing the supply of antimalarials as they shifted production focus towards more promising COVID therapeutics. In addition to these specific threats to both ITN and antimalarial coverage in Africa, there were fears that the pandemic would decrease health seeking and health systems capacity and complicate supply pipelines, leading to reductions in both widespread antimalarial access and the routine provision of ITNs even in the absence of mass campaigns.

In the midst of such speculation and concern, understanding the effect that various service disruptions might have on malaria burden became a crucial roadblock. The Malaria Atlas Project (MAP) was one of several modeling institutions tasked by the World Health Organization (WHO) and the Bill and Melinda Gates Foundation (BMGF) with circumscribing this landscape. Specifically, these policy groups were curious about the increase in malaria incidence and mortality that might be expected from service disruptions to antimalarial access, cancellation of mass ITN campaigns, reduction in capacity of routine ITN services, and combinations of these scenarios.

The full results of this modeling exercise are described in Weiss et al. [5]. Because the author of this thesis contributed modeling only to the ITN coverage analyses, it is the ITN methods and results that are emphasized here. However, because the malaria burden results showcase the importance of the ITN modeling strategy developed in Chapter 3, some emphasis is also placed on malaria incidence and mortality outcomes in the results and discussion.

4.2 Methods

The goal of this analysis, conducted in April 2020, was to estimate the impact of different ITN distribution strategies in 2020 on ITN use in the same year. Because use was the only outcome of interest here, the terms "use" and "coverage" are used interchangeably except where specified. To achieve a range of potential ITN coverage maps, the existing ITN modeling infrastructure was rerun with six different 2020 distribution scenarios. These estimates were subsequently used along with counterfactual antimalarial coverage scenarios to estimate potential impacts of COVID-related service disruption on malaria incidence and mortality in 2020.

For 2000 to 2019, the stock and flow models were run exactly as described in Chapter 3. When considering ITN distributions in 2020, modelers and WHO colleagues collaborated to consider relevant COVID-related disruptions. Here the distinction between mass and routine campaigns became important. In early 2020, many countries with mass campaigns scheduled for that year were considering postponing or canceling them. Additionally, it was anticipated that COVID-related changes in ITN supply, health-seeking, and education might reduce the capacity of routine net distribution channels such as antenatal clinics, immunization centers, and schools. Therefore, stakeholders were interested firstly in the impact of mass campaign cancellations for those countries which had them scheduled, and secondly

in the additional impact of scaling down routine distributions for all countries. In addition to altering the input data, slight changes were made to model parameters for 2020 to reduce variability in the different scenarios.

4.2.1 2020 Data

The six scenarios explored are listed in Table 4.1. Unlike earlier versions of the stock and flow model, which did not distinguish between mass and routine distributions, this analysis required separate estimates for the two distribution types in 2020. Data on planned mass ITN distributions from 26 countries were obtained from the African Leaders Malaria Alliance (ALMA) via WHO (private correspondence). The remaining 14 countries did not have mass campaigns planned for this year (i.e., the "Business as Usual" and "No Mass Campaign" scenarios were identical). See Table 4.2 for more details.

The estimation of routine distribution numbers followed two main strategies. For the 23 countries with President's Malaria Initiative (PMI) Malaria Operational Plan (MOP)s available (https://www.pmi.gov/resource-library/mops/fy-2020), projected routine distribution counts were extracted from those plans. For the remaining countries, a typical routine distribution count was estimated by taking the minimum reported distribution count from 2014 to 2018. These routine campaign counts were scaled down by 0.75, 0.5, and 0.25 for the "reduced routine" scenarios. See Table 4.2 for more details.

Details of Routine and Mass Distributions, 2020				
Country	**Mass campaign scheduled?**	**Mass Distribution Count (Millions)**	**Routine Data Source**	**Routine Distribution Count (Millions)**
Angola	No	0	PMI	2.53
Benin	Yes	9	PMI	0.75
Burkina Faso	No	0	PMI	1.7

Details of Routine and Mass Distributions, 2020, cont.				
Country	**Mass campaign scheduled?**	**Mass Distribution Count (Millions)**	**Routine Data Source**	**Routine Distribution Count (Millions)**
Burundi	Yes	6.38	Minimum	0.73
Cameroon	Yes	3	PMI	0.25
Central African Republic	Yes	0.97	Minimum	0.06
Chad	Yes	9.5	Minimum	0.38
Comoros	Yes	0.44	Minimum	0.01
Congo	No	0	Minimum	4.50e-04
Cote d'Ivoire	Yes	20.5	PMI	3.4
Dem. Rep. of Congo	Yes	24	PMI	3.6
Djibouti	No	0	Minimum	0
Equatorial Guinea	No	0	Minimum	0.01
Eritrea	Yes	1.96	Minimum	0.06
Ethiopia	Yes	6.3	PMI	0.5
Gabon	No	0	Minimum	4.60e-03
The Gambia	No	0	Minimum	0.09
Ghana	No	0	PMI	1.3
Guinea	No	0	PMI	0.47
Guinea-Bissau	Yes	1.34	Minimum	0.06
Kenya	Yes	15.8	PMI	2.37
Liberia	No	0	PMI	0.31
Madagascar	No	0	PMI	0.68
Malawi	No	0	PMI	1.2
Mali	Yes	8.17	PMI	1.63
Mauritania	Yes	1.5	Minimum	0.05
Mozambique	Yes	11.8	Minimum	1.34
Niger	Yes	8	PMI	1.83
Nigeria	Yes	16.4	PMI	1.63
Rwanda	Yes	7.5	PMI	1.19
Senegal	No	0	PMI	1.45
Sierra Leone	Yes	4.6	PMI	0.63
Somalia	Yes	1	Minimum	0.29
South Sudan	Yes	7.4	Minimum	0.45
Sudan	Yes	5.6	Minimum	2.73
Tanzania	Yes	10.2	PMI	7.16
Togo	Yes	5.97	Minimum	8.60e-03
Uganda	Yes	28	PMI	4.25
Zambia	Yes	9.37	PMI	2.09
Zimbabwe	No	0	PMI	0.8

Details of Routine and Mass Distributions, 2020, cont.				
Country	**Mass campaign scheduled?**	**Mass Distribution Count (Millions)**	**Routine Data Source**	**Routine Distribution Count (Millions)**

Table 4.2: Details of 2020 routine and mass ITN distribution numbers, by country. In the routine data source column, "Minimum" indicates that the minimum of NMCP distributions from 2014 to 2018 was used as the estimated default distribution for 2020.

This chapter presents analyses conducted in April and May of 2020, whereas the final analysis described in Chapter 3 was conducted in October 2020. The latter analysis includes surveys from five countries that were not available at the time of this work: the 2018 Cameroon DHS, the 2019 Ghana MIS, the 2018 Mali DHS, the 2018 Uganda MIS, and the Zambia 2018 DHS.

4.2.2 2020 Model

For most years, model prior distributions on net distributions are intentionally wide, since this data source is the most uncertain. Additionally, some amount of excess stock is permitted to be distributed every year to fit to survey data (see Section 3.3.3). For 2020, however, since specific scenarios were being explored, it was desired to distribute the precise number of nets recorded in the input data. Therefore, for the year 2020 only, the stock and flow prior distribution on per-capita net distributions was narrowed from $\text{Unif}(0, 0.03)$ to $\text{Unif}(0, 0.001)$, and the amount of excess stock that could be distributed was set to zero. Every other component of both the stock and flow model and geostatistical regression model remained unchanged for all six scenarios.

Even though all models were re-fit for each scenario, most model outputs stay constant. In the stock and flow model, because all changes were made after the latest survey data point, historical time series were expected to stay very similar.

Scenario	Mass Campaigns?	% of Routine Campaign	Name in Figure 4.4 and Figure 4.5
Business as Usual	Yes	100	Baseline
No Mass Campaigns	No	100	-
75% of Routine	No	75	ITN -25% (Scenario 4)
50% of Routine	No	50	ITN -50% (Scenario 5)
25% of Routine	No	25	ITN -75% (Scenario 6)
No Nets Distributed	No	0	-

Table 4.1: Scenario names and descriptions for COVID mitigation analysis.

For the regression models, the use gap regression was expected to remain identical between scenarios, while the access deviation regression was expected to differ in 2020. Because the geospatial regression shares information over both space and time, it was possible for the access deviation results for different scenarios to differ in years just before 2020 or in countries neighboring those with large differences between scenarios. Mapped nets-per-capita was not relevant to this analysis.

Once maps of ITN use from 2000 to 2020 were generated, four of the six scenarios (Table 4.1) were used to generate estimates of malaria incidence and mortality in 2020. Varying antimalarial coverage was explored in addition to varying ITN coverage [5]. Antimalarial disruption scenarios were calculated by taking the map of antimalarial coverage in 2019 as a baseline, and scaling this coverage down by 25, 50, and 75%. All antimalarial coverage and malaria burden modeling was conducted by Drs Susan Rumisha, Daniel Weiss, and Samir Bhatt.

4.3 Results

The ITN use time series from different scenarios highlight a number of phenomena that influence 2020 net coverage (Figure 4.1). First, as could be expected, countries with large mass campaigns planned in 2020 experienced the largest gap between the best- and worst- case coverage scenarios. Mauritania, Guinea-Bissau, Benin, South

Country	**Median Retention Time, This Analysis (Years)**	**Median Retention Time, Chapter 3 (Years)**
Cameroon (CMR)	3.49	3.64
Ghana (GHA)	1.78	1.87
Mali (MLI)	2.81	3.07
Uganda (UGA)	1.66	1.89
Zambia (ZMB)	1.32	1.26

Table 4.3: Median LLIN retention times in this analysis that are meaningfully different from those in Chapter 3 due to differences in survey data availability.

Sudan, Togo, and Uganda all show this effect clearly (although Benin, which had already completed its mass campaign by spring of 2020, was in no actual danger of the low coverage scenarios shown here). Second, countries that happened to have completed a mass campaign in the years immediately prior to 2020 were more resilient to COVID-related disruptions than countries that didn't. Senegal's worst-case scenario, for example, proffers higher coverage than Guinea-Bissau's, most likely due to the different phasing of three-year mass campaign cycles. Third, the "No Nets Distributed" scenario makes evident how much of an impact net retention can have on net coverage. In this scenario, Malawi and Zimbabwe arrived at similar coverage levels by the end of 2020 (11.8 and 12.5% use, respectively) despite Malawi starting from 53.3% use at the beginning of the year and Zimbabwe starting from 24.0% use. Zimbabwe's estimated median net retention time is 2.81 years in this scenario, while Malawi's is only 1.33 years.

LLIN median retention times in all scenarios were almost identical to those reported in Section 3.5.3, except for those countries in which additional survey data were added between the two analyses (Table 4.3). Even in these cases, median retention times do not differ by more than half a year.

Tanzania presents a unique example. In addition to regular mass campaigns, some Tanzanian provinces distribute large numbers of nets routinely through schools [6], such that almost as many nets were scheduled to be distributed via routine

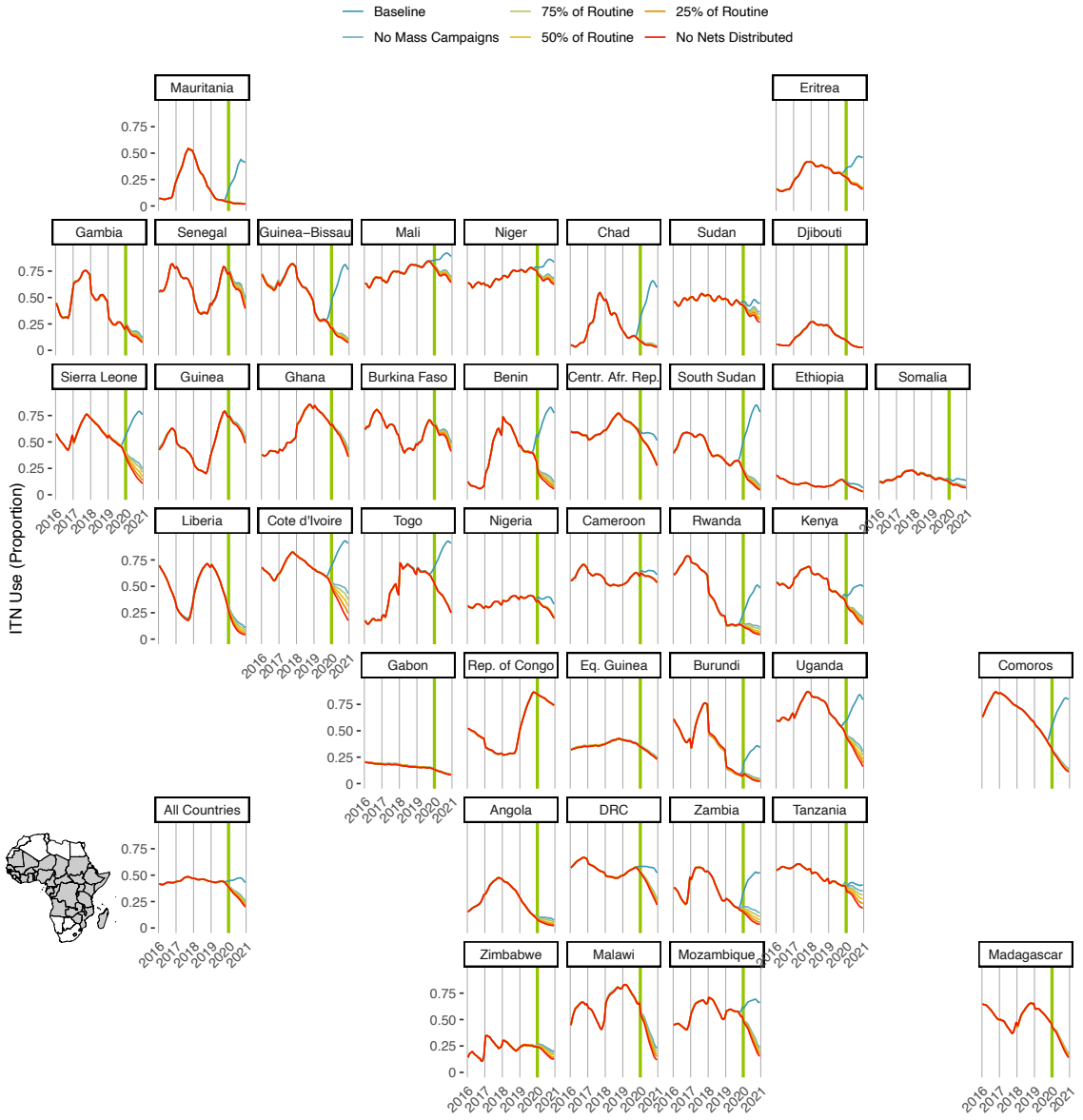


Figure 4.1: Time series comparison of Insecticide-Treated Net (ITN) use under different 2020 ITN distribution scenarios. Use is shown here as a proportion.

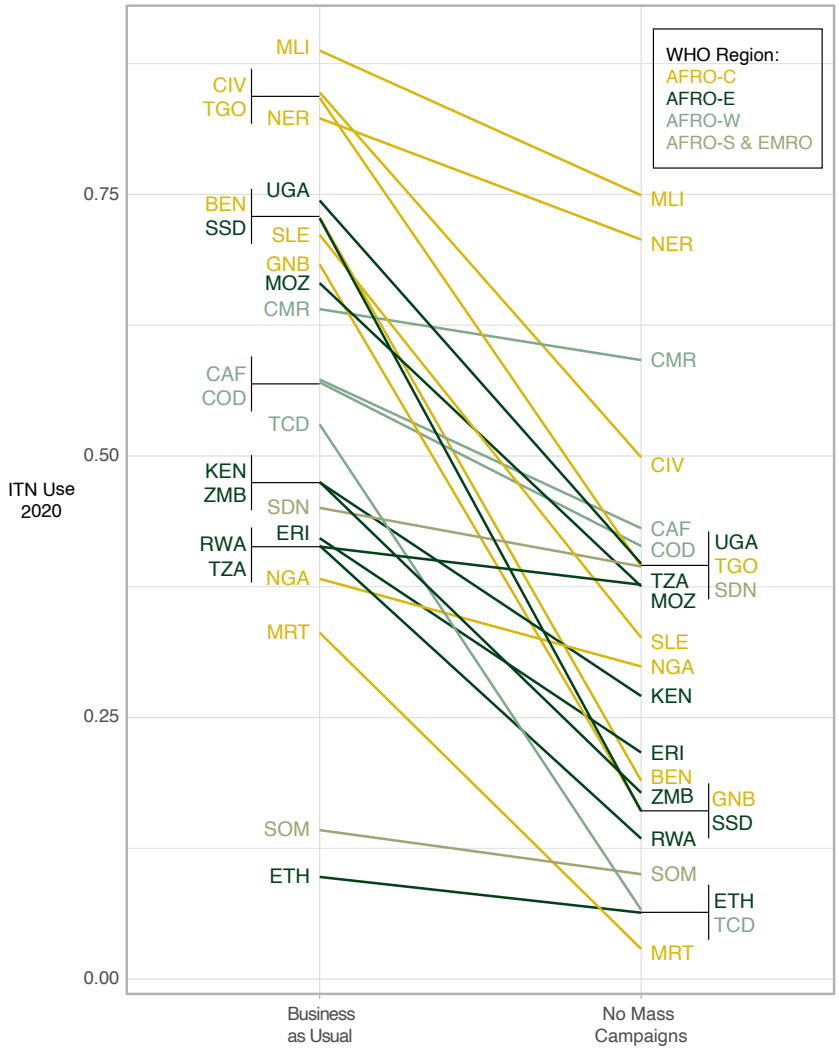


Figure 4.2: Comparison of national-level Insecticide-Treated Net (ITN) use with and without mass campaigns in 2020 for all countries with a mass campaign planned. Use is shown here as a proportion.

channels as via mass campaigns in 2020 (7.16 and 10.2 million nets, respectively). Combined with a relatively long median net retention time of 2.15 years, Tanzania showed considerable resilience to mass campaign cancellation in this exercise. Other countries that showed smaller impacts of mass campaign cancellation, such as Cameroon, Somalia, and Ethiopia, had more modest mass campaigns planned for 2020 (Figure 4.2).

As expected given that the same survey data were used for all scenarios, the spatial distribution of ITN use within a given country in 2020 changed very little

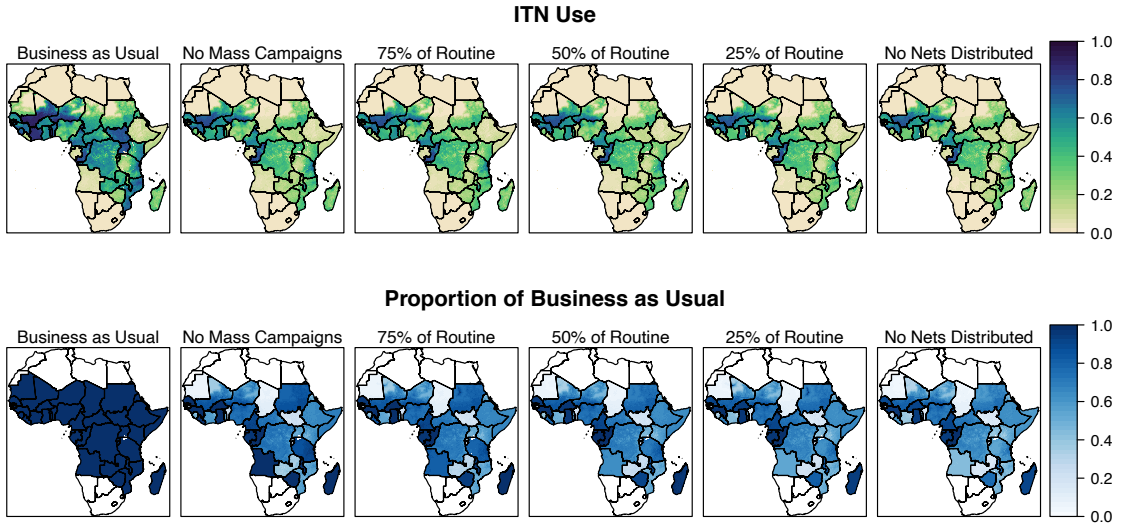


Figure 4.3: Maps of estimated Insecticide-Treated Net (ITN) use in 2020 (top), and use as a proportion of the "Business as Usual" scenario (bottom). Use is shown here as a proportion.

from scenario to scenario (Figure 4.3). The spatial view remains informative as a marker of countries deeply impacted by disruptions, such as South Sudan, compared to those more protected in 2020 such as Cameroon.

When translated into impact on malaria indicators in 2020, and compared against disruptions to antimalarial coverage, different patterns are seen for incidence and death. At the continental scale, increases in incidence are on the same order of magnitude when considering disruptions to either service (Figure 4.4). Increases in deaths, on the other hand, are much more pronounced for antimalarials than for ITNs. Drilling down to the country level reveals more heterogeneity in the relative impact on incidence of disrupting either resource, while the pattern for deaths remains consistent (Figure 4.5). One notable exception is South Sudan, which had very low antimalarial coverage at baseline and a record-setting mass campaign planned in 2020. In this setting, the impact on deaths was more pronounced for ITN service disruption than for antimalarial service disruption.

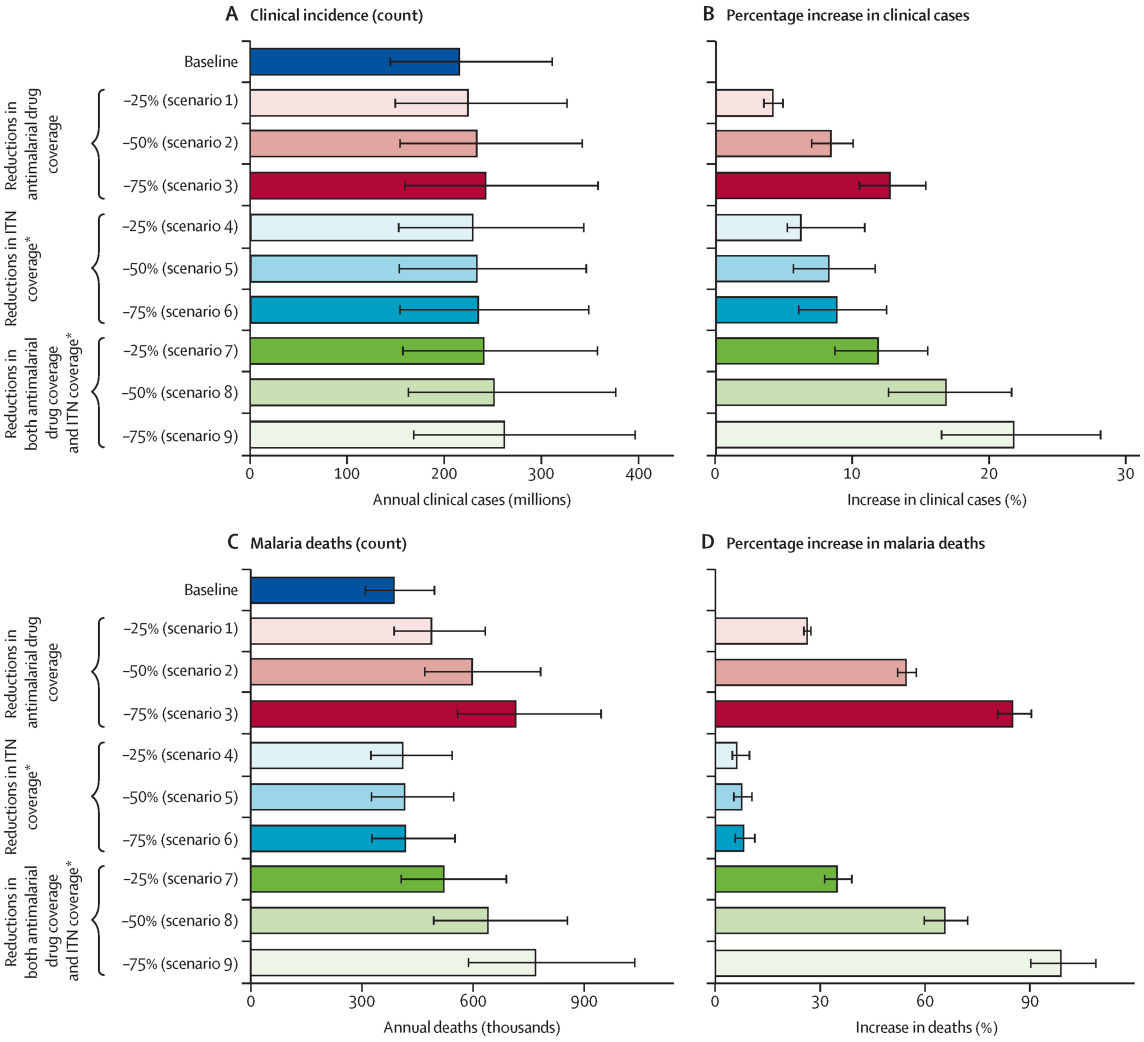


Figure 4.4: Impact of Insecticide-Treated Net (ITN) and antimalarial service disruption scenarios on malaria incidence and mortality in 2020 across all countries in this analysis. Error bars indicate 95% credible intervals. Figure generated by Daniel Weiss and Jennifer Rozier and used with permission.

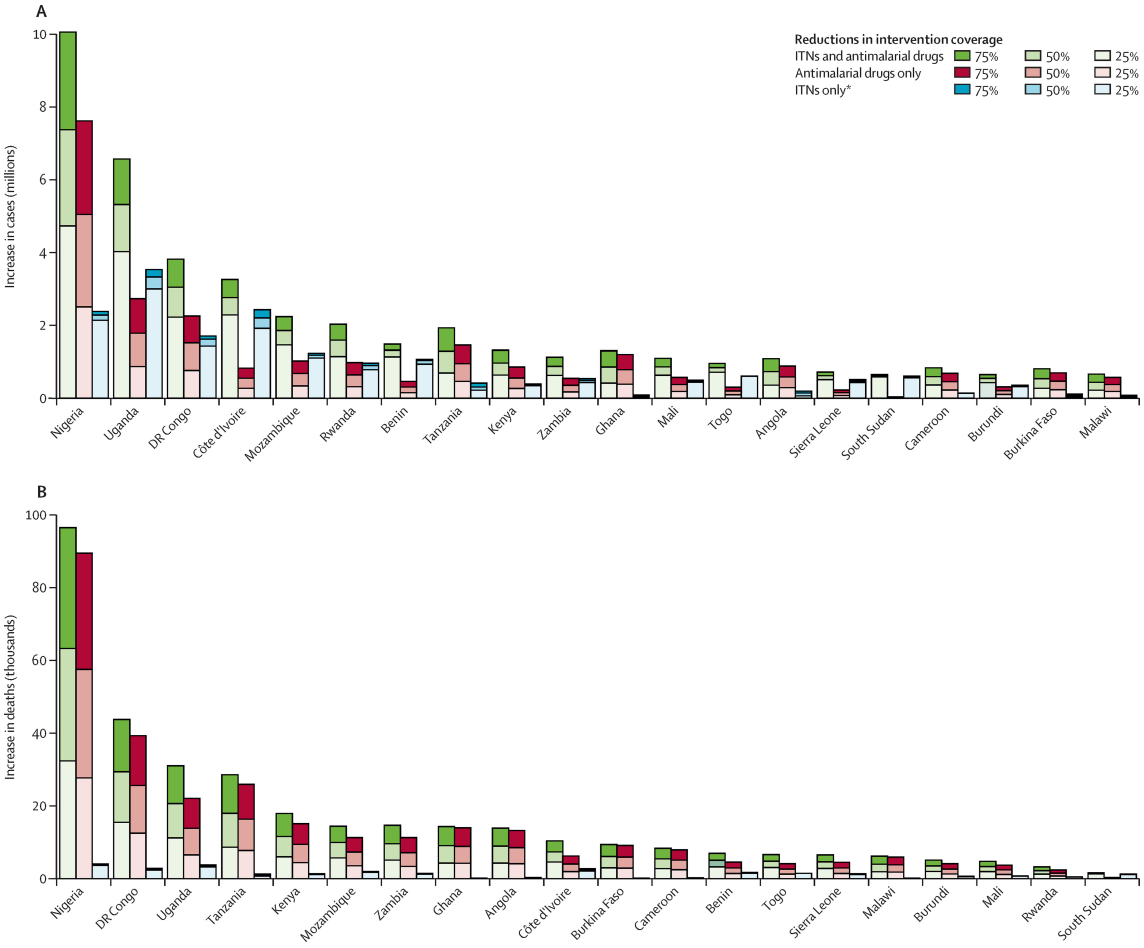


Figure 4.5: Impact of Insecticide-Treated Net (ITN) and antimalarial service disruption scenarios on malaria incidence and mortality in 2020 for the 20 most-impacted countries in this analysis. Figure generated by Daniel Weiss and Jennifer Rozier and used with permission.

4.4 Discussion

The labor-intensive process of independently modeling net distribution and retention for each country in this analysis over time was crucial to appropriately assessing how ITN coverage might change under different service disruption scenarios. Within this framework, projected ITN coverage behaved as expected in all cases. Feeding these different coverage scenarios into the malaria burden models and comparing to antimalarial service disruption scenarios, however, produced surprising results that show the importance of carefully accounting for net retention in ITN coverage projections.

The results shown in Figure 4.4 were initially startling given that an earlier analysis had suggested a much larger impact of ITNs than antimalarials on the prevention of malaria cases [7]. Closer inspection, however, revealed that the results of these two analyses were very much aligned. The apparent discrepancy can be resolved by noting two key factors: the inclusion of mortality as a metric and the short time scale of the present analysis.

Bhatt et al. [7] assessed the relative impact of ITNs, antimalarials, and indoor residual spraying on malaria incidence and prevalence from 2000 to 2015. Mortality was not considered in that analysis, and it is the mortality estimates of the present analysis which show the starkest contrast between impact of antimalarial service disruption and ITN campaign disruption. This makes perfect sense given that the primary purpose of antimalarials is to prevent death among children who have already acquired malaria. When antimalarials are available and used at a sufficiently high level, they may also reduce incidence by preventing infectious parasites from being transmitted onward, but this is a secondary goal. ITNs, conversely, work primarily as a preventive tool, and their impact on mortality comes purely as a result of averting infections. Therefore, the larger impact of antimalarial disruption on

mortality is unsurprising. Why, then, does antimalarial disruption still have a comparable or greater impact than ITN disruption on clinical incidence? The argument given above suggests the opposite should be true. The explanation for this result can be found in the short time scale of this analysis, and the fact that ITN coverage has inertia in a way that antimalarial coverage does not.

The goal of this analysis was to assess the impact of malaria service disruptions in 2020 on malaria burden *in 2020*. Antimalarials are consumable products with a relatively short lifespan– they can be used only once and supplies disappear relatively quickly in the absence of regular re-stocking. ITNs, by contrast, are reusable products designed to have a lifespan of several years at least. The model to estimate malaria incidence is statistical, not mechanistic, but it was still able to reflect this relationship because the antimalarial covariate was calculated using a simple scaling factor, while the ITN covariate was calculated based on a country-specific mechanistic model that takes net lifespan into account. Thus, while the antimalarial disruption scenarios modeled a sudden, stark decrease in coverage of an important protective covariate, even the most pessimistic ITN scenarios were cushioned by the presence of nets already in the community. This protective effect of extant nets explains the relatively minor impact of ITN service disruption on malaria incidence in 2020. By focusing on such a short time scale, however, this analysis risks obfuscating a key message: that halting ITN distributions in 2020 would have a potentially devastating impact *in future years*. Countries that distributed nets in 2019 were most protected against service disruptions in 2020. Not distributing in 2020 would leave countries deeply vulnerable in future years, especially if new shocks arose to threaten net distributions yet again.

Results from this analysis were included in a WHO report in April 2020 [3]. This and other evidence was used by policymakers to successfully advocate for proceeding with mass ITN campaigns in all countries which had scheduled them, as well as

negotiating solutions for continued delivery of antimalarials. The impact of these decisions has yet to be assessed, but may be substantial [8].

This analysis has several limitations. While the values for planned mass campaign ITN distributions come directly from the relevant sources, data for typical routine net distribution numbers were harder to come by, and in some cases had to be assumed from prior data. Therefore, routine distribution counts may be less reliable than mass campaign counts. This analysis was run by re-fitting the full time series model with different ITN distribution data, which left open the possibility that different scenarios might have different model fits. While there are some instances of coverage deviations in 2019 (in South Sudan and Zambia, for example), these did not enormously affect the overall results. A more principled version of this analysis would have kept the historical time series unchanged and projected forward for 2020 only, but time did not permit the creation of this new model infrastructure. Limitations of the non-ITN portions of this work are listed in Weiss et al. [5].

The ITN model was not developed with the specific intention of projecting potential coverage scenarios into the future, yet its strategy for mechanistically estimating country-specific time series within a geospatial estimation framework proved instrumental in generating realistic coverage estimates for COVID-related service disruptions. This modeling infrastructure was able to quickly and effectively produce maps of projected ITN coverage in a time of considerable urgency and with provable policy impact. Without the built-in flexibility of a country-specific mechanistic model, this framework would have not been able to provide specific and timely feedback to policymakers on this crucial topic. Future model improvements should keep this use case in mind should additional crises require similar types of analysis.

Bibliography

1. Cohen, J. M. *et al.* Malaria resurgence: a systematic review and assessment of its causes. *Malaria Journal 2012 11:1* **11,** 1–17. ISSN: 1475-2875. https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-11-122 (2012).
2. WHO. *Tailoring malaria interventions in the COVID-19 response Global Malaria Programme* tech. rep. (). http://www.who.int/malaria.
3. WHO. *The potential impact of health service disruptions on the burden of malaria* tech. rep. 23.4. 2020 (2020), 1–38. https://www.who.int/publications/i/item/the-potential-impact-of-health-service-disruptions-on-the-burden-of-malaria.
4. Gabler, E. & Keller, M. H. *Prescriptions Surged as Trump Praised Drugs in Coronavirus Fight - The New York Times* https://www.nytimes.com/2020/04/25/us/coronavirus-trump-chloroquine-hydroxychloroquine.html (2021).
5. Weiss, D. J. *et al.* Indirect effects of the COVID-19 pandemic on malaria intervention coverage, morbidity, and mortality in Africa: a geospatial modelling analysis. *The Lancet Infectious Diseases* (2020).
6. Massue, D. J. *et al.* Durability of Olyset campaign nets distributed between 2009 and 2011 in eight districts of Tanzania. *Malaria Journal* **15,** 176. ISSN: 14752875. http://www.malariajournal.com/content/15/1/176 (2016).
7. Bhatt, S. *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* **526,** 207–211. ISSN: 14764687 (2015).
8. Roll Back Malaria. *Campaign Dashboard — RBM Partnership* https://endmalaria.org/dashboard/campaign-dashboard (2021).

Conclusion

> Now that you know what you know, how will you use it to inform your personal and professional practices?
>
> Sandra Styers
> *Indigenous and Decolonizing Studies in Education*

Overview

This document has described the development and refinement of two frameworks for combining mechanistic and geospatial models in malaria research in a manner that can easily and explicitly inform large-scale policy decisions. Both of these modeling infrastructures are methodologically novel, programatically valuable, and accessible to other researchers in the field.

The first framework, focused on malaria transmission, uses machine learning tools and spatial data to identify key archetypes for a mechanistic transmission model, then utilizes spatial data again to place mechanistic model results in a global context. This strategy is flexible, customizable, and computationally efficient (Chapter 1). It uses an original quantitative strategy to solve the computational problems of continent-level intervention impact assessment without loss of spatial detail. Once archetypes and representative sites have been selected, arbitrarily de-

tailed intervention simulations can be run to generate continent-wide estimates in a matter of minutes or hours. In the use cases shown here, machine learning methods detected signals in 27-dimensional data across 985,206 pixels on the continent of Africa to identify seasonal and entomological archetypes that align well with established seasonality maps. In contrast to previous methods, which assessed rainfall and entomology separately, this strategy was able to jointly explore rainfall, temperature, and entomological landscapes to group locations across multiple parameters of interest. Next, three-year simulations were run for 152 intervention packages to generate detailed maps of estimated intervention impact across the continent.

The closest published analogue to the archetypes framework is that described in Walker et al. [1]. In this setting, four rainfall classes, four entomological classes, and 18 transmission intensity classes were combined with 343 interventions for a total of 98,784 scenarios. In contrast, the six-class archetypes setting run for 18 transmission intensities and 343 interventions would have required only 37,044 scenarios. As an added computational benefit, because a single simulation run can capture multiple locations, these would in practice have required only 6,174 simulations.

As described in Chapter 2, results from this framework were used in 2018 and 2019 to inform major international eradication policy decisions made by both the WHO Strategic Advisory Group on Malaria Eradication and the Lancet Commission on Malaria Eradication. Once the framework was developed, new intervention packages could be added and visualized with a turnaround time of days. This feature has proven useful on multiple occasions since. For example, when Bill and Melinda Gates Foundation (BMGF) stakeholders requested burden reduction estimates under a range of different Insecticide-Treated Net (ITN) retention time scenarios in summer 2021, the archetypes framework was able to rapidly implement and test these intervention packages without the overhead of setting up a brand new simulation infrastructure.

The second framework, focused on insecticide-treated net coverage, uses a mechanistic commodity tracking model as the backbone for a spatiotemporal regression model in order to appropriately capture different countries' unique histories of net procurement, distribution, and retention. This work is a complete restructuring and expansion of an earlier ITN model including an entirely rewritten and publicly-available code base, addition of previously unused data, correction of bugs and structural issues, and addition of new features including maps of nets per capita and ITN use rate. In addition to these technical innovations, this project made time series maps of four ITN coverage metrics from 2000 to 2020 publicly available for the first time. The core model described in Chapter 3 contributed to the 2019 and 2020 World Malaria Reports and has been published in *Nature Communications*, where it has been accessed over 2,000 times since its publication in mid-June 2021. In addition to the continued used of this framework in annual ITN coverage and malaria burden estimates for WHO, this work sparked interest from the Clinton Health Access Initiative (CHAI). Colleagues at Malaria Atlas Project (MAP) have been working with CHAI representatives to generate commodity forecasting dashboards based on ITN framework results. The ITN retention analysis for BMGF stakeholders described in the previous paragraph was originally founded on results from the *Nature Communications* paper. An emergency modification of the ITN model in light of the emerging COVID-19 pandemic played a crucial role in informing WHO policy around prioritizing ITN distributions in 2020 (Chapter 4).

Both of these modeling frameworks showcase the value of combining the flexibility and versatility of mechanistic modeling with the statistical rigor and geographical tools of geospatial modeling.

Future Work

Each of these frameworks has a lengthy list of downstream projects. I am actively working with stakeholders at BMGF to use the archetypes framework to assess how net durability and retention might impact malaria burden in different settings. I plan to collaborate with colleagues at Institute for Disease Modeling (IDM) to consider ways of incorporating historic intervention time series into the archetype creation process, a necessary step before the framework is ready to compare against real-world results in data-rich settings. Colleagues at MAP have used ITN model results to craft a commodities planning dashboard for stakeholders at CHAI that may also find users in national programs and elsewhere. I plan to make a careful descriptive comparison of bednet coverage and malaria transmission intensity, possibly as a precursor to more technical work assessing the complex causal pathways in this space. There is ample room for the ITN model to grow, allowing for the incorporation of insecticide resistance data as well as other markers of net efficacy and impact.

Concluding Thoughts

From my childhood collecting animal encyclopedias through the completion of my Masters in Public Health in 2016, I uncritically loved science and data. I thought that both of these concepts were objective and immovable, and that if more people would just use them the right way, we could save the world. I did not deeply interrogate who decided what the "right" way was, who I thought was doing the saving, and who I imagined needed to be saved.

Since beginning my DPhil in 2017, in parallel to my educations about malaria transmission and statistics and computer science, I have been pursuing an uncomfortable education in history. I have learned terrible things about those infrastruc-

tures of western science that I grew up loving: how they have systematically appropriated or erased indigenous ways of knowing and being [2, 3], how they have been crafted to pathologize and exploit colonized peoples [4–6], and how they still work today to exclude nonwhite and nonmale voices from power and agency, in global health and elsewhere [7–10]. I have glimpsed the ways in which my own work, situated as it is within neocolonial power structures, contributes to these injustices. There is a great deal in this thesis I am proud of, but I am not proud of that.

I regret that I lacked the self-awareness, bravery, and time it would have taken to base my DPhil on work that challenged this colonial status quo. I am, however, delighted and energized by the sea change in mindset that has been gaining velocity in global health over the past decade. Thousands of scientists from historically excluded countries and identities are working brilliantly to claim the resources owed to them and redefine the field to truly center and serve their communities. In my future work, I commit to contributing to this movement to the best of my abilities: to center the work and priorities of researchers and policymakers from malaria-endemic countries, to avoid or reject projects that lack representation from the communities under study, and to use what voice and power I have to advocate for equity. I invite and encourage anyone reading these words to hold me accountable to these claims.

It has been a joy and a privilege to devote the last four years to this work. I am thrilled that this is only the beginning.

Bibliography

1. Walker, P. G. T., Griffin, J. T., Ferguson, N. M. & Ghani, A. C. Estimating the most efficient allocation of interventions to achieve reductions in P.falciparum malaria burden and transmission in Africa. *The Lancet Global Health,* 1–11. ISSN: 2214109X. http://dx.doi.org/10.1016/S2214-109X(16)30073-0 (2015).
2. Tuhiwai Smith, L. *Decolonizing Methodologies* Second. ISBN: 978-1-84813-950-3. https://www.bloomsbury.com/us/decolonizing-methodologies-9781786998132 (Bloomsbury Publishing, London, 2012).
3. *Making and Unmaking Public Health in Africa* (eds Prince, R. J. & Marsland, R.) ISBN: 978-0-8214-2058-4. https://www.ohioswallow.com/book/Making+and+Unmaking+Public+Health+in+Africa (Ohio University Press, 2014).
4. Krieger, N. *Epidemiology and the People's Health* ISBN: 978-0-19-934842-8 (Oxford University Press, 2011).
5. Kendi, I. X. *Stamped from the Beginning: The Definitive History of Racist Ideas in America* ISBN: 9781568584638. www.boldtypebooks.org (Bold Type Books, New York, 2016).
6. *Making Kin Not Population* (eds Clarke, A. C. & Haraway, D.) ISBN: 978-0-9966355-6-1. https://press.uchicago.edu/ucp/books/book/distributed/M/bo28583407.html (Prickly Paradigm Press, Chicago, 2018).
7. Tuck, E. & Yang, K. W. *Decolonization is not a metaphor* tech. rep. 1 (2012), 1–40. https://jps.library.utoronto.ca/index.php/des/article/view/18630.
8. Erondu, N. A. *et al.* Open letter to international funders of science and development in Africa. *Nature Medicine,* 1–2. ISSN: 1078-8956. http://www.nature.com/articles/s41591-021-01307-8 (2021).
9. Abimbola, S. *The foreign gaze: Authorship in academic global health* 2019. http://gh.bmj.com/.
10. Abimbola, S. *et al. Addressing power asymmetries in global health: Imperatives in the wake of the COVID-19 pandemic* 2021. https://doi.org/10.1371/journal.pmed.1003604.g001.

Attributions and Publications

Chapter 1

The work in this chapter is entirely mine, and the text is in preparation for a submission to *Malaria Journal*. I will be first author of seven, with the other authors having contributed mentorship, editorial guidance, and/or software support.

Chapter 2

As noted in the text, this chapter describes work conducted jointly by myself, Dr. Peter Gething, and Dr. Samir Bhatt. I conducted all mechanistic intervention impact modeling, while Drs. Gething and Bhatt contributed the geospatial analysis assessing the impact of secular trends on malaria burden.

While the language of this chapter is entirely original, the analysis described contributed to a WHO report available at https://www.who.int/publications/i/item/9789240003675 and a publication in *The Lancet Commissions*, available at https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(19)31139-0/fulltext.

I am listed in the acknowledgements of the WHO report and descriptions of the results of Chapter 2 cover pages 33-43 of the report, but I did not directly write any

part of the manuscript.

I am listed as fourth author on the *Lancet Commissions* paper, of which Section 2 describes the analysis conducted in Chapter 2. Drs. Peter Gething and Samir Bhatt wrote the first draft of this paper section, while Samir Bhatt and I were the primary editors and respondents to reviewers during the peer review process. Peter Gething and I wrote Appendix 1 of the Supplementary Materials of this report, which describe the methods of Chapter 2 in detail.

Chapter 3

The work in this chapter is entirely mine. I lightly edited this text into a first-author manuscript, which was accepted without revisions in *Nature Communications*. It was also listed as a featured article on the *Nature* Public Health page (https://www.nature.com/articles/s41467-021-23707-7). Co-authors Drs. Peter Gething, Samir Bhatt, Caitlin Bever, Deirdre Hollingsworth, and Hannah Koenker were actively involved in guiding and developing this project both analytically, theoretically, and in crafting messaging for the paper. Co-authors Camilo Vargas-Ruiz, Dr. Harry Gibson, and Joseph Harris contributed software support. The remaining 15 co-authors read and gave feedback on the final manuscript.

Chapter 4

As noted in the text, this chapter describes work conducted jointly by myself, Drs. Dan Weiss, Peter Gething, Samir Bhatt, Susan Rumisha, and others. I contributed all ITN coverage analyses, while my co-authors contributed analysis of other intervention coverage and malaria burden.

While the language of this chapter is entirely original, the analysis described contributed to a WHO report available at https://www.who.int/publications/i/item/9789240004641 and a publication in *Lancet Infectious Diseases*, available at https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30700-3/fulltext.

I am listed in the acknowledgements of the WHO report, but I did not directly write any part of the manuscript and ITN coverage results are not specifically described.

I am listed as second author on the *Lancet Infectious Diseases* paper. Drs. Daniel Weiss, Peter Gething, and I were primary authors, editors, and respondents to reviewers during the peer review process.

Acknowledgements

All avenues to love, I realized,
were acts of imagination.

YZ Chin
Edge Case

I have so many people to thank.

Thanks to Pete Gething for years of generative conversations and good advice. Thank you especially, Pete, for giving me the space and permission to waste time, especially on this ITN project that I'm sure was not what you had in mind for me. Thanks to Sam Bhatt for being endlessly generous with his time and deeply gracious to a student who was picking his past work apart with a scalpel. I'll pass it on, Sam.

Thanks to Caitlin Bever for her unwavering patience, support, and attention to detail. Thanks to Deirdre Hollingsworth for her warmth and strategic insights. Thanks to Jaline Gerardin for the fastest copy-editing I have ever seen, an encyclopedic knowledge of both EMOD and malaria literature, and for teaching me how to structure a paper. Thanks to all three of these women for embodying the type of scientist, mentor, and colleague I would like to be.

Thanks to Philip Welkhoff for being my first global health friend, for almost a decade of mentorship, and for showing me a model of leadership that centers empathy and abolishes condescension.

Thanks to too many colleagues and mentors to count, but to name just a few: Katherine Battle, Dan Weiss, Camilo Vargas-Ruiz, Susan Rumisha, Punam Amratia, Harry Gibson, Joe Harris, Monique Ambrose, Jon Russell, Prashanth Selvaraj, Andre-Lin Ouedraogo, Josh Suresh, Shirley Leung, Dan Bridenbecker, Svetlana Titova, Edward Wenger, Josh Proctor, Letitia Onyango, Hannah Koenker, Beatriz Galatas, Abdisalan Noor, Gerry Killeen, Sam Kiware, Jenn Gardy, Bruno Moonen, Justin Cohen, and many more. You have made this work such a pleasure.

Thanks to the friends who have gifted their love, affection, and time to me, in ways large and small: Dan, Dan, Kate, Alex, Dave, Dave, Hannah, Caitlin, Maggie, Enzo, Os, Becca, Mike, Jon, Travis, and more. Thanks especially to Dan Goes for personally carrying my heart through the days when it was too heavy for me to hold.

Thanks to my mother Elena for teaching me to trust my instincts, question authority, and cook for myself. Thanks to my sister Clara for being my biggest champion, lifelong hero, and personal stylist. I am undefined without both of you. Thanks to Greg for helping me with my math homework, finding the best walks through the woods, and constantly showing me new ways of being family. Thanks to Kennedy for my childhood and for shaping my view of fierceness and justice and play. Thank you, Papà, for life and humor and a sharp tongue.

Thanks to my dance teachers Kiné, Kaitlin, Christin, Natascha, and Jaret. Thanks to Liberty and the gang, especially Max and Dalan. Thanks to the Cowley Road Brewdog, the Oxford Museum of Natural History, Peloton Espresso, and Gail's Bakery. Thanks to the unceded Duwamish land upon which I write these words, and the people that land was stolen from.

Finally, thanks to Hugo: Floppy dog, gentle bear, swamp monster, Mr. Handsome. If there is joy in this document, it's because of you.