Measuring inequities in health service delivery and their impact on health outcomes in low- to middle-income countries

Adrien Allorant

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Robert C. Reiner, Jr., Chair

Nancy Puttkammer

David Pigott

Program Authorized to Offer Degree:

Global Health

©Copyright 2021
Adrien Allorant

University of Washington

Abstract

Measuring inequities in health service delivery and their impact on health outcomes in low- to middle-income countries

Adrien Allorant

Chair of the Supervisory Committee:
Associate Professor Robert C. Reiner, Jr.
Health Metrics Sciences

Scientific evidence increasingly indicates that poor-quality health systems will slow down the tremendous improvements in health outcomes observed in low- to middle-income countries (LMICs) over the past several decades. However, much of this evidence focuses on country-level assessment of health systems performance, derived from measures of excess mortality. This approach does not enable examinations of within-country inequities in access to quality health services. Yet geographical factors, such as rurality, road networks, and elevation, affect accessibility to health services. Moreover, the inputs that condition quality care - drug supply, testing services, qualified health workers - are not uniformly distributed in space. This suggests that access to quality health services varies sub-nationally. Furthermore, there is scant literature on the effect of quality of care improvements on reduced disease burden, such as under-five mortality. Sub-national inequities in health service delivery can materialize as the failure to reach certain groups, populations or geographies in need for care and/or failure to provide care of sufficient quality. This dissertation aimed at presenting statistical methods that could be used to leverage available survey data and estimate metrics of health services delivery, at the scale of programmatic implementation, to identify priority-settings and interpret the effects of interventions and policies over time.

In the first chapter, *Where are the missing tuberculosis cases in Bangladesh? Finding*

gaps in routine surveillance activities in Bangladesh by linking prevalence survey and case notification data, we developed a Bayesian spatial model to jointly analyze two independent data sources: the 2015-16 national tuberculosis sero-prevalence survey and the tuberculosis cases notified yearly to the National Tuberculosis Control Program in Bangladesh. Our study estimated important indicators of the tuberculosis epidemic, including the prevalence-to-notification ratio, a key metric of surveillance activities performance, at the second administrative division level (districts). Our results indicated that despite considerable progress in case notifications and treatment coverage in Bangladesh, substantial within-country inequalities on key indicators of the TB epidemic were still observed. We performed a counter-factual analysis to quantify the potential number of people who develop tuberculosis every year, and could be diagnosed and notified if sub-national disparities were resolved, and found that most people missing from care each year were concentrated in selected districts. Enhanced routine surveillance activities along with active case-finding strategies will be key to reach, diagnose, and treat people with tuberculosis in these areas that likely face barriers in accessing testing and treatment services. Sub-national estimates of prevalence and prevalence-to-notification ratios can guide these efforts, help develop tailored strategies, and ensure that no population is left behind in the fight against tuberculosis.

The second chapter, *What do we know about readiness and process quality of care in low-to middle-income countries? A statistical framework to measure sub-national inequalities in readiness and process quality metrics using health facility surveys, in Senegal, Kenya, and Tanzania*, introduced a statistical framework to characterize levels and trends in readiness and process quality of sick-child care metrics, from health facility surveys, at sub-national resolutions. Large standardized health facility surveys provide the most consistent and comparable description of the level at which patients interact with the health system; yet, analyses of these data sources to date have focused on cross-sectional, national studies, overlooking temporal trends and sub-national inequities. We developed Bayesian hierarchical models to

analyze all publicly available survey data in a given country, and explored spatio-temporal variations in metrics of readiness and process quality. We applied our framework to three countries with several rounds of facility surveys- Senegal, Kenya, and Tanzania. Our modelling approach allowed us to supplement direct survey measurements with spatio-temporal smoothing and auxiliary information from spatially resolved covariates (such as health workers density or indicators of urban development) to improve the precision of our estimates for areas with sparse data. We validated our model through out-of-sample predictions. We found evidence of substantial spatio-temporal heterogeneity in health services quality metrics. Improving health system quality requires new methods that enhance the measurements of actionable quality metrics, at a sub-national resolution that can help inform programmatic decision-making. Our study proposed a statistical framework to derive such metrics from health facility surveys.

In our third chapter, *Metrics of quality-adjusted health service coverage for sick children and child mortality in Senegal, Kenya and Tanzania: an ecological regression study*, we introduced statistical methods to estimate quality-adjusted coverage metrics, at sub-national resolution and over time, using household and health facility surveys. Quality-adjusted coverage for sick-child care, a measure of the proportion of children with symptoms of pneumonia, diarrhea, or malaria, that were diagnosed and treated at a health facility according to evidence-based guidelines, is a critical metric to identify and resolve gaps in the availability, readiness and quality of child health services. We applied our framework to three countries with several rounds of household and facility surveys- Senegal, Kenya, and Tanzania, and found overall low levels of coverage, with substantial heterogeneity across sub-national units. We estimated four different metrics of quality-adjusted coverage, corresponding to more or less concise metrics of readiness and process quality of sick-child care, and compared their performance in predicting under-five mortality rates, over time, and at sub-national levels, using an ecological regression framework. Our model included area-level estimates of risk

factors (such as diarrhea and lower-respiratory infection incidence) and preventive interventions (such as immunization coverage) to account for differences between regions. We found a negative association between quality-adjusted coverage and under-five mortality rates at the ecological-level, and no evidence of lower predictive performance of more concise quality-adjusted coverage metrics. As quality-adjusted coverage becomes a more common metric of health system performance, our study provides important insights for decision makers. Our analysis finds that spatio-temporal estimates uncovered substantial sub-national disparities. This finding can be used to tailor data collection efforts, as more concise constructs to measure readiness and process quality of care can be performed.

TABLE OF CONTENTS

Page

List of Figures iii

List of Tables v

Glossary vii

Chapter 1: Specific Aims 1

Chapter 2: Where are the missing tuberculosis cases in Bangladesh? Finding gaps in routine surveillance activities in Bangladesh by linking prevalence survey and case notification data 3

Abstract 3

Background 5

Methods 7

Results 10

Discussion 14

Conclusion 18

Acknowledgements 18

Chapter 3: What do we know about readiness and process quality of care in low- to middle-income countries? A statistical framework to measure sub-national inequalities in readiness and process quality metrics using health facility surveys, in Senegal, Kenya, and Tanzania 19

Abstract 19

Introduction 21

Methods 23

Results 28

Discussion 39

Conclusion . . . 44

Chapter 4: Metrics of quality-adjusted health-service coverage for sick children and child mortality in Senegal, Kenya and Tanzania: an ecological regression study . . . 48

Abstract . . . 48

Introduction . . . 50

Methods . . . 52

Results . . . 59

Discussion . . . 66

Conclusion . . . 70

Chapter 5: Conclusions . . . 71

Summary Aims 1-3 . . . 71

Future works . . . 73

Bibliography . . . 76

Appendix A: Appendix: Chapter 2 . . . 88

A.1 Supplementary Data . . . 88

A.2 Supplementary Methods . . . 88

A.3 Supplementary Results . . . 91

Appendix B: Appendix: Chapter 3 . . . 100

B.1 Supplementary data . . . 100

B.2 Supplementary Methods . . . 102

Appendix C: Appendix: Chapter 4 . . . 114

C.1 Supplementary Data . . . 114

C.2 Supplementary Methods . . . 116

C.3 Supplementary Results . . . 117

LIST OF FIGURES

Figure Number | Page

2.1 Distribution of surveyed clusters across Bangladesh's divisions and observed prevalence of TB (per 100,000 population) at these clusters 8
2.2 Estimated TB mean prevalence (per 100,000) (a), and 97.5th (b) and 2.5th (c) percentiles 11
2.3 Estimated mean number of people with TB (a), and 97.5th (b) and 2.5th (c) percentiles 12
2.4 Estimated mean TB prevalence-to-notification ratio (a), and 97.5th (b) and 2.5th (c) percentiles 13
2.5 Estimated mean TB additional cases that could be detected if every district achieved at least the national prevalence-to-notification ratio (a), and 97.5th (b) and 2.5th (c) percentiles 14

3.1 Estimated readiness metrics (general Service Readiness Index and child Service Readiness Index) across sub-national units in Senegal, Kenya and Tanzania, in 2010 and 2020 33
3.2 Mean estimates of adherence to IMCI protocols (a,c,e), and length of the 95% interval (b,d,f) estimates, in Senegal (a,b), Kenya (c,d), and Tanzania (e,f) in 2020 34
3.3 Facilities' general SRI in Senegalese departments, in 2013 (a) and 2019 (b) . 38
3.4 Prevalence of fever history assessment during sick-child visits in Senegal, in the departments where malaria is endemic (left panel) or seasonal (right), during the dry or rainy seasons 46
3.5 Prevalence of fever history assessment during sick-child visits in Tanzania, in the regions where malaria is endemic (left panel) or seasonal (right), during the dry or rainy seasons 47

4.1 Health-service cascade coverage for sick-child care 54
4.2 Metrics of quality-adjusted coverage over time in Senegal, Kenya and Tanzania 61
4.3 Mean estimates of quality-adjusted coverage metric 3 in 2020, in Senegal (a), Kenya (b) and Tanzania (c) 62

4.4 Service contact and receipt of intervention coverage in Senegal, Kenya and Tanzania, by sub-national units in 2010 and 2020 63

A.1 Design-based (Horvitz-Thompson) estimates of TB prevalence by district, 2015–2016 Districts left blank (labelled "NA") are the districts that did not contain any cluster sampled during the survey in 2015–2016. 91

A.2 Predicted TB median prevalence (per 100,000) with the best (a), the smoothing (b) and the CPO (c) model 94

A.3 Predicted median prevalence of household overcrowding in Bangladesh between 2007 and 2015 95

A.4 Relapse cases per 1,000 TB cases in Bangladesh in 2016 96

B.1 Direct survey estimates coverage by model-based uncertainty interval, in Senegal. 105

B.2 Direct survey estimates contained in the model-based 95% credible intervals, in Kenya. 106

B.3 Direct survey estimates contained in the model-based 95% credible intervals, in Tanzania. 107

B.4 IMCI diagnostic protocols (a) and General SRI (b) estimates, in Senegal in 2010 and 2020 110

B.5 Child IMCI diagnostic protocols (a) and child SRI (b) estimates, in Kenya in 2000 and 2020 111

B.6 General SRI estimates, in Tanzania between 2005 and 2020 112

B.7 Joint distribution of the two process quality of care metrics in Senegal, Kenya and Tanzania, over the study period. Breaks were created at the 33rd and 66th percentiles for the two metrics. 113

C.1 Service and receipt of intervention coverage quantiles in 2010 (a) and 2020 (b) at the second administrative level in Senegal. Service and receipt of intervention coverage quantiles in 2010 (c) and 2020 (d) at the first administrative level in Kenya. Service and receipt of intervention coverage quantiles in 2010 (e) and 2020 (f) at the first administrative level in Tanzania. 117

C.2 Temporal random effects modelled as a first order random walk 118

C.3 Spatially correlated random effects 119

C.4 Posterior densities of the fixed effects for quality-adjusted coverage under the four models (on the transformed-logit scale) 119

C.5 Comparing posterior mean estimates of quality-adjusted coverage between metrics 4 and 1, in Senegal, Kenya, and Tanzania 120

LIST OF TABLES

Table Number Page

3.1 List of surveys available in Senegal, Kenya and Tanzania 25
3.2 List of tracer items and diagnostic protocols used to derive readiness and process quality of care 26
3.3 List of covariates considered in fitting equation (3.2) 28
3.4 Selected model for each readiness and process quality metric, and variance decomposition. Columns $\sigma^2_{S_i}$, $\sigma^2_{\gamma t}$, $\sigma^2_{\alpha_t}$, $\sigma^2_{\delta_{it}}$, $\sigma^2_{\nu_s}$, and $X_{it}\boldsymbol{\beta}$ indicate the percentage of total variance explained by each component. 29
3.5 Performances of the models, assessed using mean squared errors (MSE) and spatial leave-one-out predictions for the four metrics, across Senegal, Kenya and Tanzania 31
3.6 Summary inequality measures, across sub-national units of Senegal, Kenya and Tanzania 36

4.1 List of surveys available in Senegal, Kenya and Tanzania 53
4.2 Components of quality-adjusted health-service coverage for sick-child care . . 55
4.3 List of tracer items and diagnostic protocols used to derive readiness and process quality of care, by approach 57
4.4 Posterior means of the fixed effect coefficients for the four models 65
4.5 Performances of the four models using the four different quality-adjusted metrics, assessed using mean squared errors (MSE) and spatial leave-one-out predictions of U5MR in the 118 sub-national units across Senegal, Kenya and Tanzania 66

A.1 Covariate data sources 89
A.2 Model comparison using different set of covariates) 92
A.3 Predictors' effect estimates 93
A.4 Complete list of district-level prevalence estimates per 100,000 inhabitants (1/3) 97
A.5 Complete list of district-level prevalence estimates per 100,000 inhabitants (2/3) 98
A.6 Complete list of district-level prevalence estimates per 100,000 inhabitants (3/3) 99

B.1 Characteristics of the health facilities sampled in the SPA, by year in Senegal, Kenya and Tanzania 101
B.2 Characteristics of the health facilities sampled in the SDI, by year in Senegal, Kenya and Tanzania 108
B.3 Models considered to estimate quality metrics over time and space 109

C.1 List of covariates considered to model under-fiver mortality rates by subnational units 115
C.2 Models considered to estimate the association between quality-adjusted coverage metric and under-five mortality rates, by subnational units and over time 116

GLOSSARY

LMICS: Low- to middle-income countries

TB: Tuberculosis

WHO: World Health Organization

UN: United Nations

SDGS: Sustainable Development Goals

GBD: Global Burden of Disease

NTP: National Tuberculosis control Program

USAID: United States Agency for International Development

DHS: Demographic and Health Survey

WB: World Bank

SPA: Service Provision Assessment

SDI: Service Delivery Indicators

SARA: Service Availability and Readiness Assessment

IMCI: Integrated Management of Childhood Illness

UI: Uncertainty Interval

ACKNOWLEDGMENTS

This dissertation is the culmination of many years of academic, professional and life experiences, in which several people have played a key role in various ways, and I would like to acknowledge some of them. First and foremost, I want to acknowledge my Seattle friends, who were there during the highs and lows of the past four years that, taken together, were some of the happiest in my life so far- Jamal, Joey, Neal, Aisha, the Beacon Hill house, the NOLA crew and our boy king, CC friends. These lovely and loving people made me feel at home so far away from home. Second, I must thank my dissertation committee - Bobby Reiner, David Pigott, Nancy Puttkammer, Joe Dieleman, and Jon Wakefield, for giving me the space and freedom to define my research interests and providing support and feedback once I had figured out what I wanted to research. Special thanks to Bobby for listening to my ideas for two-plus years and making sure that they would turn into a body of scientific work. Thanks to David for always stretching my thinking; I always left our conversations with tons of new ideas on my research. Special thanks to Nancy for bringing me to health services research, and for being such a tremendous mentor from our work in Haiti to the chapters of this dissertation. Thanks to Joe for being such a personable and supportive mentor- from the first to the last days of this program; I have felt so lucky to be mentored by someone who deeply cared about my academic and non-academic life. Thank to Jon for being the best GSR I could have dreamt of- an expertise in survey/spatial statistics only paralleled by an expertise in European football and rugby. Third, I want to thank my cohort-mates, the PhD students in Global Health - Metrics and Implementation Sciences alike - for their kindness and the inspiration that procured their work, research and lived experiences. Special thanks to my classmates Stephanie, Claire, Joey, Andre, Beatrice, Barbara, Rima,

Becky, Meredith, Deanna, Kristen, Elizabeth, Quinhas, Yoshi, Nami and Sarah. I also want to thank Gregoire, who played such an important role in bringing me to Seattle and to the field of Global Health. Finally, I would like to thank my family. I am so grateful for the love and support I have received from my grandparents, my parents and my siblings throughout the years. Their dedication to knowledge, public service, and caring for others directly inspires my life path. Lastly, I want to thank my partner, Alexandrea, for being an endless source of happiness, by finding beauty and joy in everything, turning each challenge into a fun activity, and reminding me of what truly matters. As June Jordan wrote, Love is lifeforce.

Chapter 1

SPECIFIC AIMS

High-quality health systems are needed to expand the health gains in infectious diseases and child health experienced by LMICs in the past several decades [1]. It is estimated that 8.6 million deaths that occur every year in LMICs are amenable to healthcare, including 2 million deaths from just two conditions - tuberculosis and neonatal conditions [2]. Measuring healthcare accessibility and quality is key to accountability and action; yet few LMICs have data systems and institutions dedicated to monitor service delivery [3]. In their absence, quality measurements are restrained to indicators of inputs (e.g. supply of drugs, number of beds per 1,000 inhabitants) - which are poorly correlated with health outcomes - or mortality-based metrics (inferring health systems performance from amenable mortality [4]) - which are not suited to identify specific gaps in the cascade of care. Improving health system quality will require new methods that enhance the measurements of accessibility and quality of health care [5], and work to disaggregate the results for vulnerable populations and geographies.

The overarching goal of this dissertation was to generate metrics of healthcare access and quality, at a sub-national resolution that could help inform programmatic decision-making and inequality reduction. My specific aim was to develop methods and metrics best suited for data available in low-resource settings, mainly standardized surveys, to address three challenges that impede the measurements of inequities in health service delivery: (1) the gaps in routine surveillance activities that result in 41% of people developing TB every year to be missed from care, (2) the absence of disaggregated metrics of readiness and process quality of care to assess sub-national inequities, and (3) the lack of evidence directly linking metrics of readiness and process quality to health outcomes, such as child mortality.

In high-TB-burden countries, such as Bangladesh, areas with lower access to healthcare tend to have larger numbers of people developing TB every year, but are less likely to detect them, leading to high numbers of undiagnosed incident TB cases [6]. In these priority settings, we found that comparing sub-national prevalence estimates with case notification rates has the potential to identify areas where routine surveillance gaps are the largest.

Recent works have documented significant variations in the service readiness and the quality of primary care of health facilities in Senegal, Kenya and Tanzania [7]. The Service Provision Assessment (SPA) and the Service Delivery Indicators (SDI) surveys collect critical indicators of health facilities' capacity to provide essential services, such as child health care, and could be leveraged to assess spatio-temporal variations in access to quality care.

Previously produced high-resolution estimates of under-five mortality rates across 46 countries in Africa from 2000 to 2017, had highlighted substantial sub-national heterogeneity in child mortality [8, 9]. While several studies had tried to explain this spatial variability in using individual, household and community-level characteristics [10], there was a great need to quantify the impact of access to quality child health care on U5MR.

Using data from (1) a national sero-prevalence survey in Bangladesh, (2) two large standardized health facility surveys, and (3) international household surveys, we pursued three specific aims:

- **Specific Aim 1**: Examined geographical patterns of tuberculosis prevalence and case notification rates in Bangladesh to identify gaps in routine surveillance activities.

- **Specific Aim 2**: Evaluated readiness and process quality of sick-child care metrics, in Senegal, Kenya and Tanzania, sub-nationally and over time.

- **Specific Aim 3**: Explored the association between quality-adjusted coverage for sick-child care and child mortality in Senegal, Kenya and Tanzania, over time, at the ecological level.

Chapter 2

WHERE ARE THE MISSING TUBERCULOSIS CASES IN BANGLADESH? FINDING GAPS IN ROUTINE SURVEILLANCE ACTIVITIES IN BANGLADESH BY LINKING PREVALENCE SURVEY AND CASE NOTIFICATION DATA

Abstract

Background: Tuberculosis (TB) has been the leading cause of death from a single infectious pathogen globally since 2014. To reduce this burden, drivers preventing people from accessing care, including gaps in diagnosis and treatment initiation, must be better addressed. Fine-scale estimates of TB prevalence and case notification can be combined to guide priority-setting for strengthening routine surveillance activities in high-burden countries. In this study, we produce policy-relevant estimates of the TB epidemic at the second administrative unit (district) in Bangladesh.

Methods: We used a Bayesian spatial framework and the cross-sectional National TB Prevalence survey from 2015–2016 in Bangladesh to estimate prevalence by districts. TB case notification rates were used to calculate district-level prevalence-to-notification ratio, a key metric of under-diagnosis and under-reporting. We conducted a counterfactual analysis to estimate the number of additional TB cases that could be notified if each district reached at least the national prevalence-to-notification ratio.

Results: TB prevalence rates were highest in the northeastern districts and ranged from 160 cases per 100,000 (95% uncertainty interval 80–310) in Jashore to 840 (690–1020) in Sunamganj. The absolute number of prevalent cases was highest in the northeastern districts of Sunamganj, Sylhet, and Netrokona and in the densely populated districts of Dhaka and Chattogram. Despite moderate prevalence rates, Rajshahi and Dhaka divisions presented the

highest prevalence-to-notification ratios, due to low case notification. Resolving sub-national disparities in case detection could lead to 26,500 additional TB cases (8,500–79,400) notified every year.

Conclusion: This study is the first to produce and map sub-national estimates of TB prevalence and prevalence-to-notification ratios in a high-burden setting. In the absence of reliable local indicators on TB incidence, sub-national TB prevalence estimates are essential to target prevention and treatment efforts, while prevalence-to-notification ratios can support needs assessments for routine surveillance. Reaching TB cases currently missing from care will be key to ending the TB epidemic.

Keywords: Tuberculosis; Tuberculosis prevalence survey; Case notification; National Tuberculosis Control Program; SDG 3; spatial analysis; geospatial modeling; survey methods

Background

Tuberculosis (TB), with an estimated 1.2 million deaths worldwide, was the leading cause of death from a single infectious pathogen in 2019 [11]. Despite ambitious global goals set by the World Health Organization (WHO) and the United Nations (UN) to reduce the global burden of the TB epidemic, an estimated 10 million people developed TB in 2019. WHO's End TB Strategy and UN's Sustainable Development Goals (SDGs) target 3.3 aim to end the global TB epidemic, with targets to reduce TB deaths by 95% and new cases by 90% between 2015 and 2035 [12, 13]. Achieving these targets will require an unprecedented effort to close the gaps in the TB care cascade, by which only 71% of people who develop TB every year are diagnosed and initiate treatment.

To track progress towards these targets, WHO's Taskforce on TB Impact Measurement [14] and the Institute for Health Metrics and Evaluation [15] have developed national models to estimate key indicators of the TB epidemic (such as incidence, prevalence, and mortality rates) that would otherwise be challenging to measure directly, particularly in low- to middle-income countries (LMICs). TB incidence, which cannot be measured directly, is derived from routine case notification, inventory studies, or, where available, national prevalence surveys [16]. In priority settings, where under-reporting and under-diagnosis due to poor access to health services can lead to a wide discrepancy between routine case notification and the actual number of incident TB cases [6], surveys that measure TB prevalence are invaluable to assess progress towards WHO's and UN's targets.

Bangladesh is a high-burden country for TB according to WHO. The Global Burden of Diseases, Injuries, and Risk Factors (GBD) Study estimated 216,000 incident cases (95% uncertainty interval [UI] 184,400–250,200) and 29,100 deaths (UI 22,400–40,000) among HIV-negative people in 2019 in Bangladesh. With an estimated treatment coverage of 81%, approximately one in every five TB cases in Bangladesh is missing from care every year. Nationally representative TB prevalence surveys were conducted in Bangladesh in 2007–2009 and 2015–2016 and found prevalence of smear-positive tuberculosis of 79.4 per 100,000 (95%

Confidence Interval [CI] 47.1–133.8) and 113 per 100,000 (CI 87–139) in persons aged 15 years and older, respectively [17, 18]. The 2015–2016 survey, which used both smear microscopy and Xpert MTB/RIF additionally found an all-form TB prevalence of 287 per 100,000 (CI 244–330) in persons aged 15 years and older. These two surveys used different methodologies and their results are therefore not directly comparable. Both surveys revealed significant differences in prevalence by age, gender, socioeconomic status, and rural/urban status, however, suggesting substantial within-country variation in burden. An increasing number of spatial analyses of TB prevalence [19, 20, 21, 22], incidence [23], or mortality [24] in high-burden countries have also showed high within-country variation.

In addition, prior investigations of the spatial distribution of TB in Bangladesh include two recent studies from the KIT Royal Tropical Institute relying on sub-national case notification data collected by Bangladesh NTP [25, 26]. These studies both supply evidence of extensive gaps in TB notifications of certain districts (second-level administrative divisions) of Bangladesh and conclude that the case notification rate is an unreliable proxy for TB incidence. Relying solely on notification data can lead to misleading results, as areas with limited access to health care tend to have a larger number of underlying TB cases and are less likely to detect them [27]. In the absence of sub-national measurements of TB incidence, local estimates of TB prevalence can help inform prevention and treatment efforts towards areas with high TB burden. Additionally, comparing sub-national prevalence estimates with case notification rates can help identify areas where the gap between the estimated number of TB cases and the number of people diagnosed and notified is the largest.

In this study, we illustrate how the growing number of TB prevalence surveys can be used to supply sub-national estimates of prevalence in Bangladesh. We adapted an existing Bayesian spatial model [28, 29] to analyse data from the National TB Prevalence Survey (2015–2016) and produce district-level estimates and uncertainty intervals accounting for the complex survey design, and calibrated the data to national estimates from the GBD study [30]. There are several methodological challenges associated with estimating TB prevalence sub-nationally. First, the sampling design, a stratified multi-stage cluster sampling, needs to be accounted

for in the analysis. Second, TB is relatively rare in select districts in Bangladesh, where low case counts can lead to potentially unstable estimates. In this study, we address these challenges by using a Bayesian spatial hierarchical model to analyse Bangladesh's 2015–2016 national survey and produce estimates of prevalence for pulmonary TB in those 15 years and older at fine geographical scales. Additionally, we use case notification data collected by Bangladesh's NTP to estimate district-level prevalence-to-notification ratio, a key metric of the gaps in detection and reporting of new cases. Finally, we conduct a sub-national counterfactual analysis, which sets the national prevalence-to-notification ratio as the benchmark for every district, and calculated the number of additional TB cases that could be notified by resolving sub-national inequalities. Achieving the End TB Strategy and SDG 3.3 will require resources and interventions targeting locations with large numbers of TB cases that are undiagnosed or are lacking treatment coverage. This study shows how local estimates of TB prevalence and prevalence-to-notification ratio can be combined to find areas in greatest need of enhanced access to diagnosis and treatment services.

Methods

Data

Bangladesh's National TB Prevalence Survey was cross-sectional and used a multi-stage cluster sampling method [18]. Data were collected across Bangladesh (figure 2.1) from any person aged 15 years and older, living in one of the 125 sampled clusters (46 urban and 79 rural), between March 2015 and April 2016. Case definition included Xpert MTB/RIF and/or culture-confirmed cases; smear-positive subjects not confirmed by Xpert MTB/RIF were excluded for the estimation of prevalence. There were 98,710 participants in the study, including 20,594 screening positive to either symptom or chest X-ray; two sputum specimens were collected from each of the 20,010 individuals. A total of 278 bacteriologically confirmed cases were considered as study cases, including 108 smear-positive and 170 smear-negative cases. Specific details about the survey complex design and findings are presented extensively

elsewhere [18]. The TB notification data reported by each of the 64 districts of Bangladesh were collected from the NTP. New and relapse pulmonary TB cases reported by district in 2016 were used to calculate a district-level prevalence-to-notification ratio.

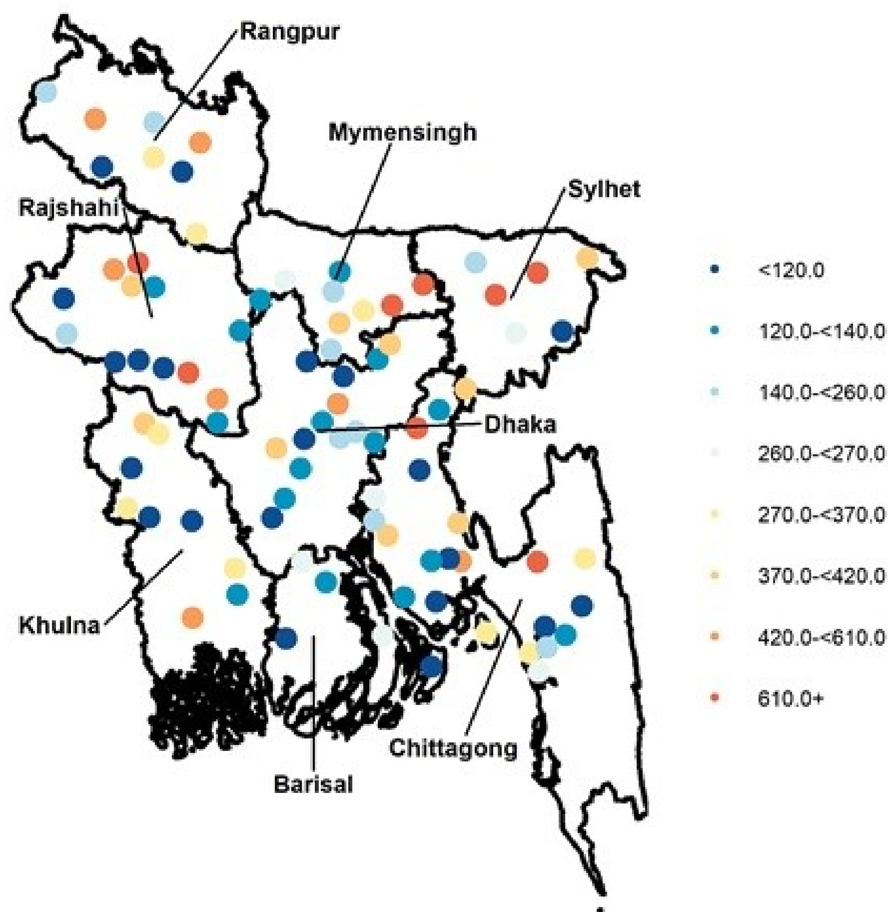


Figure 2.1: Distribution of surveyed clusters across Bangladesh's divisions and observed prevalence of TB (per 100,000 population) at these clusters

Statistical analysis

TB prevalence was estimated using a small area estimation approach, which is a spatial extension of the classic Fay and Herriot approach [31]. First, the survey design was accounted

for by directly calculating the Horvitz-Thompson estimator of TB prevalence for each district using the design weights (see supplementary methods A.2). Second, the logit transformation of the true district prevalence was modelled as a linear function of covariates and spatially structured random effects. Out of the 64 districts of Bangladesh, 19 contain no sampled clusters. In these locations, our hierarchical approach drew strength from spatially adjacent districts and modelled covariates to predict TB prevalence. Direct estimates and design-based variance estimates were computed using the *survey* package in **R**. We fit the Bayesian hierarchical models using the Integrated Nested Laplace Approximation [32] and the *R-INLA* package [33]. We obtained a subset of all included covariates by checking for multi-collinearity using the variance inflation factor with a threshold of 5, and compared the models consisting of different combinations of these selected covariates, using four selection procedures (the residual mean squared error [RMSE], the deviance information criteria [DIC], the Watanabe-Akaike information criteria [WAIC], and the conditional predictive ordinate [CPO]). In the absence of consensus on a single criterion, we select the best-performing model across the criteria (see supplementary methods A.2.1), and present the estimates from the second and third best performing models as supplementary results (see supplementary results A.2). Predicted prevalence estimates were divided by the total number of pulmonary cases in those aged 15 years and older notified for the corresponding district in 2016. We conducted a counterfactual analysis in which districts with a prevalence-to-notification ratios above the national average prevalence-to-notification ratio were set to the national average. We used the counterfactual analysis to estimate the number of additional TB cases that could be notified by the NTP every year, if every district reduced its prevalence-to-notification ratio at least the national average. The national prevalence-to-notification ratio, which was 2.8 in 2016, was used as a benchmark. By using the average prevalence-to-notification ratio nationally, we showed the potential impact of reducing sub-national heterogeneity in surveillance and TB-control activities on TB case detection.

Results

District-level estimates of TB prevalence revealed a clear spatial pattern in TB prevalence, with higher rates in northern and eastern districts than south-western districts. North-eastern districts displayed the highest TB prevalence rates: over 80% larger than 287 per 100,000, the national average prevalence in the population 15 years and older in Bangladesh (figure 2.a). Specifically, Sunamganj, with 840 cases per 100,000 (UI 690–1020), Netrokona with 640 cases per 100,000 (580–690), and Sylhet with 520 cases per 100,000 (370–710), were the three districts with the highest prevalence rates nationally (see supplementary table A.4 for the complete list of district-level prevalence estimates). Predicted TB prevalence overall varied significantly by district: Sunamganj, Netrokona, and Sylhet, the three districts with prevalence rates in the 95th percentile nationally, had prevalence rates three times higher than districts in the 5th percentile, such as Natore, with 180 cases (90–340) per 100,000, and Jashore, with 160 cases (80–310) per 100,000. Uncertainty around these estimates was high in some districts; however, five districts – Sunamganj, Netrokona, Sylhet, Gazipur, and Khulna – displayed prevalence rates significantly higher than the national average (figure 2.2). The estimated number of TB cases exhibits a more complex spatial distribution than prevalence rates. The districts with the highest number of TB cases were a mix of geographies with high prevalence rates and highly populated districts (figure 2.3). The former includes Sunamganj, Sylhet, and Netrokona, accounting for the third, fifth, and eighth largest numbers of prevalent cases, with 11,800 cases (UI 9,700–14,400), 10,788 cases (7,800–14,900) and 8,500 cases (7,800–9,300), respectively. However, the highest estimated number of TB cases was in Dhaka, the densest district of Bangladesh, with 17,300 prevalent cases (11,400–26,000). Chattogram, the district that houses the second largest city in the country, accounted for the second highest number of TB prevalent cases, 12,300 (7,900–18,800). Nevertheless, there is clustering of high TB burden in the north-eastern districts in Bangladesh when compared to the south-western parts of the country. Maps (b) and (c) on figures 2.2 and 2.3, displaying uncertainties, are useful to clearly show the

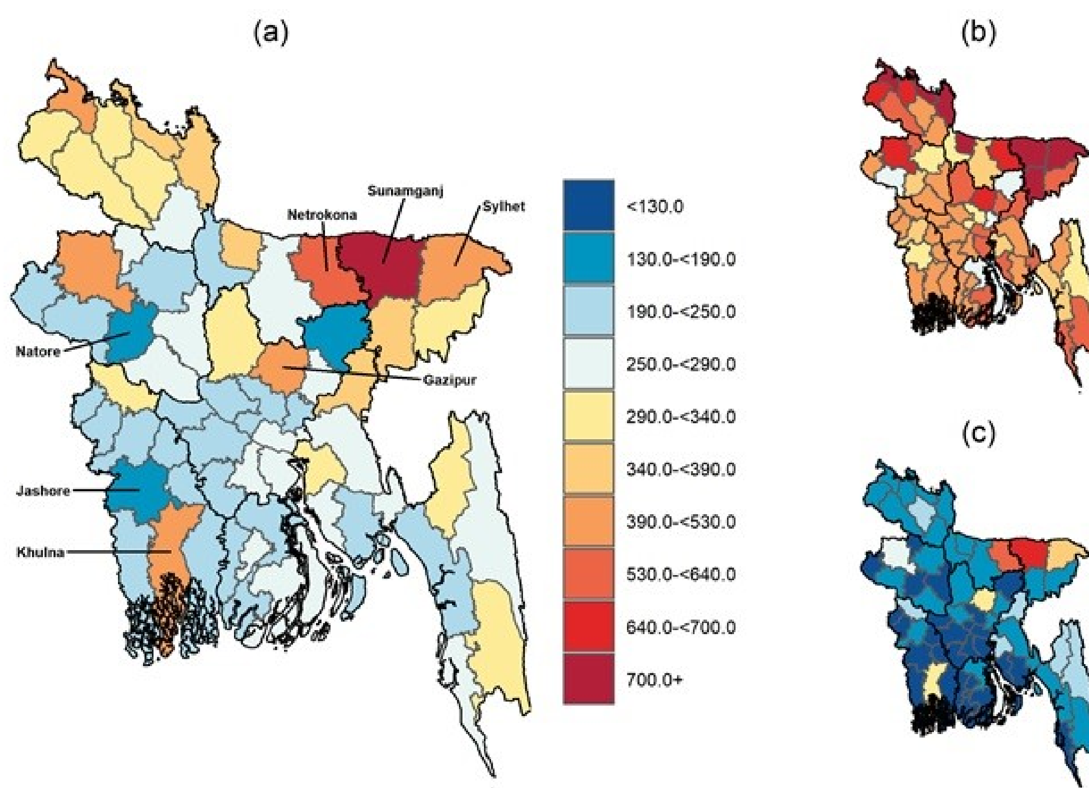


Figure 2.2: Estimated TB mean prevalence (per 100,000) (a), and 97.5th (b) and 2.5th (c) percentiles

significantly disproportionate TB burden that affects the eastern region of Bangladesh. The geographical patterns of TB prevalence-to-notification ratio (figure 2.4) mirror some of the features of TB prevalence rates; prevalence-to-notification ratios are larger than the national average, 2.8, in most districts of the northern divisions. For instance, Naogaon (7.9 UI 5.3–12.0]), Rajbari (6.7 [3.2–14.1]), Faridpur (6.1 [2.9–12.8]), and Netrokona (5.9 [5.4–6.5]) display prevalence-to-notification ratios in the 90th percentile nationally. Additionally, figure 2.4 supplies supplemental insights into the understanding of the TB epidemic in Bangladesh. Despite low TB prevalence rates, most districts of the Rajshahi and Dhaka divisions showed a large prevalence-to-notification ratio. For instance, the districts of Rajbari and Faridpur

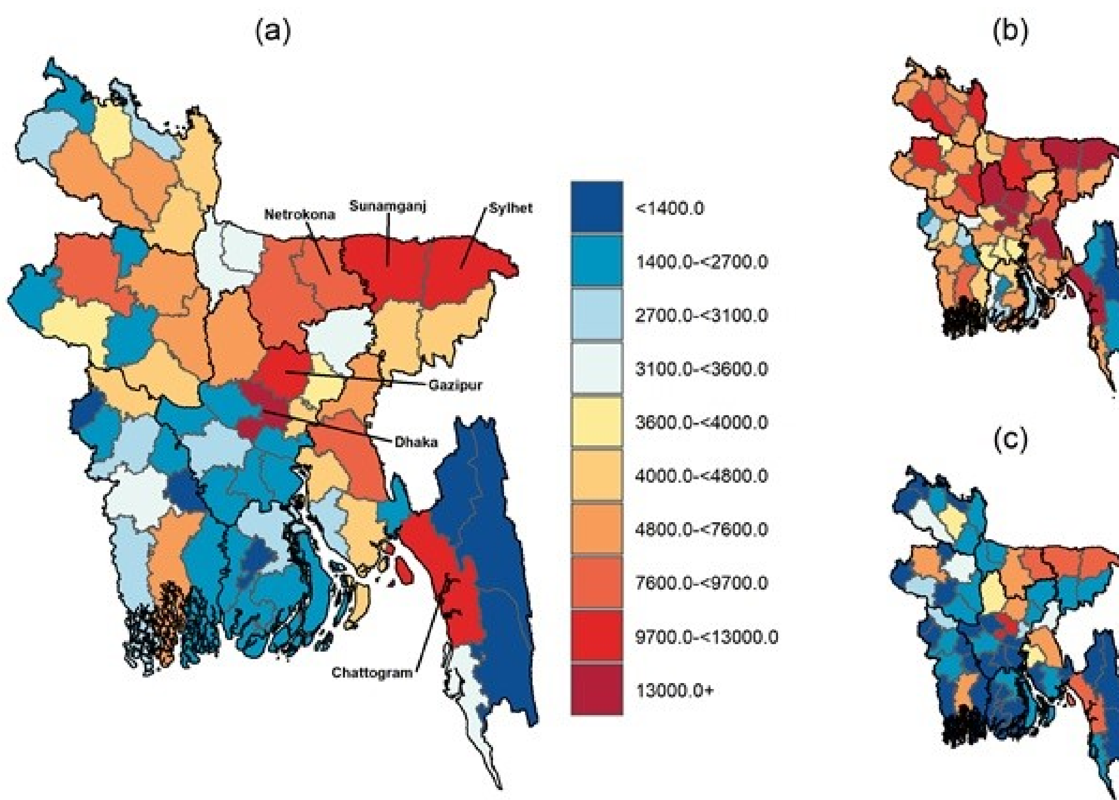


Figure 2.3: Estimated mean number of people with TB (a), and 97.5th (b) and 2.5th (c) percentiles

had the second and third largest prevalence-to-notification ratios nationally, despite moderate TB prevalence rates, 240 per 100,000 (UI 110–500) and 240 per 100,000 (120–510), respectively. These rates can be explained by strikingly low case notification rates in 2016, 31 and 36 per 100,000, in Rajbari and Faridpur, respectively. More generally, 12 out of the 13 districts in the bottom 20% of case notification rates nationally were in Rajshahi and Dhaka divisions, which seems to account for the cluster of high prevalence-to-notification ratios in the Southwest districts of these two divisions (figure 2.4). With a country mean prevalence-to-notification ratio of 2.8, the NTP can detect around one-third of all potential TB cases nationally (9). However, our results suggest substantial sub-national variations in

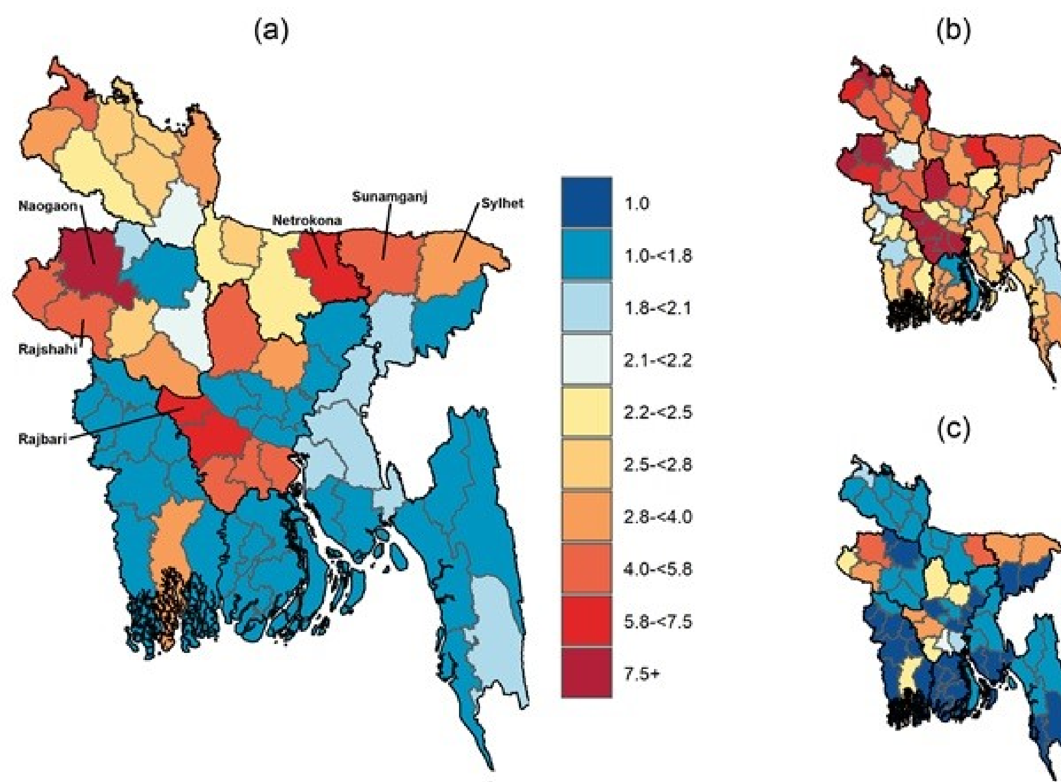


Figure 2.4: Estimated mean TB prevalence-to-notification ratio (a), and 97.5th (b) and 2.5th (c) percentiles

case-detection performance in Bangladesh, and even within the same divisions. For instance, prevalence-to-notification ratios ranged from 1.4 to 7.9 in the districts of Rajshahi division. Assuming the NTP could strengthen routine surveillance and case-detection strategies such that every district reduced its prevalence-to-notification ratio to at least the national average of 2.8, we estimated that an additional 30,800 individuals (UI 10,000–84,500) with TB could be diagnosed, notified, and started on life-saving TB treatment. The districts of Naogaon, Sunamganj, Netrokona, and Sylhet accounted for more than half of this reservoir of potential cases, with 15,900 additional notified cases (7,900–27,400) (figure 2.5).

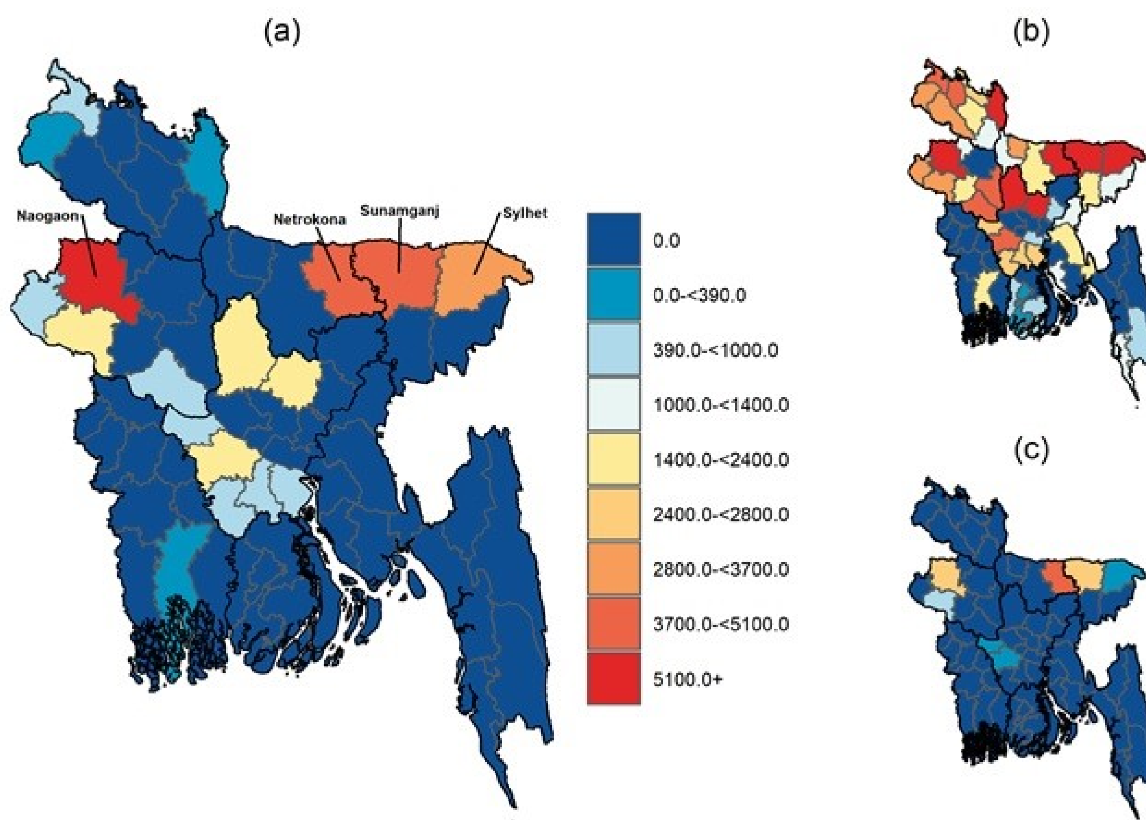


Figure 2.5: Estimated mean TB additional cases that could be detected if every district achieved at least the national prevalence-to-notification ratio (a), and 97.5th (b) and 2.5th (c) percentiles

Discussion

This study is, to our knowledge, the first nationwide study investigating within-country variations in TB prevalence while accounting for a complex survey design. Our findings suggest that despite considerable progress made nationally in reducing TB incidence, significant inequalities in the TB prevalence rates, the number of TB prevalent cases, and the prevalence-to-notification ratio persist both between and within the eight divisions of Bangladesh. First, prevalence estimates were highest in the northern and north-eastern dis-

tricts of Bangladesh, with rates more than double the national average. In the absence of reliable local indicators on TB incidence, sub-national TB prevalence estimates are essential to target prevention efforts in areas with higher ongoing transmission. Intensive TB-control interventions targeted at high burden areas could effectively reduce the high prevalence rates highlighted in this study. Interestingly, these north-eastern districts additionally display a higher-than-average proportion of relapse among new cases (supplementary results A.4), an indicator of programme performance in starting and keeping TB patients on treatment. Geographically distant from Dhaka, the populations living in these mostly rural districts may experience an increased time and financial burden for taking part in the demanding 6–12-month TB treatment course dispensed at health facilities [34]. Second, the highest number of prevalent cases were found in these same districts, along with the most populated areas, including the districts of Dhaka and Chattogram, where the two largest cities of Bangladesh are located. In high-density urban living, high contact rates, household overcrowding, and the influx of young adults migrating from rural areas or other countries can contribute to sustaining the TB epidemics [35]. Active case finding and screening contacts of TB patients are promising strategies to improve TB case detection and reduce transmission, especially in urban settings [36]. Understanding the sub-national distribution of TB cases in Bangladesh is also crucial to informing policies and ensuring the provision of quality TB services and treatment in locations where they are most needed.

Finally, our joint study of prevalence and notification data expanded the body of evidence of significant under-reporting and under-diagnosis of TB cases in Bangladesh. The results of our analysis indeed confirm the existence of a local cluster of low case notification found by the study from the KIT Royal Tropical Institute [25]. In this study, Rood and colleagues lacked sub-national measures of TB prevalence, and instead used an indicator of poverty to standardise case notification rates and identify indications of under-detection and under-diagnosis. Drawing from both prevalence and case notification data, our study provides stronger evidence of under-reporting and under-diagnosis in both the cluster of districts in the divisions of Rajshahi and Dhaka, where TB prevalence rates are moderate, but case no-

tification is strikingly low, and in the north-eastern districts, where TB prevalence rates are high. Each of these “missing” cases represents a missed opportunity to notify people developing TB, which in turn hinders progress towards ending the TB epidemic. The prevalence-to-notification ratio displayed the largest spatial disparities of the three metrics considered in this analysis, with a range of 1.1 to 7.9 between districts; eight districts displayed a prevalence-to-notification ratio more than double the national average. Additionally, spatial variations in prevalence-to-notification ratios were significantly larger than the variations by gender or age groups reported by Bangladesh NTP. The prevalence-to-notification ratio is a key metric to find gaps in identification and reporting of TB cases. These substantial subnational variations in prevalence-to-notification ratio suggest that, despite steady progress in case notification nationally, large numbers of people who develop TB are missed every year by testing services. Our findings are directly relevant for decision-makers in determining priority districts with large prevalence-to-notification ratios, which call for strengthening access to high-quality diagnostic services. In high-risk areas, strategies to identify people with active disease can supplement routine surveillance activities to enhance case detection [37].

There are several limitations to this study. First, while our approach enabled us to depict the complex picture of the heterogeneous spatial distribution of TB across Bangladesh, it does not provide further information on temporal trends. While the first national prevalence survey in Bangladesh was conducted in 2007–2009, the methodology used then, sputum samples collected from every respondent without screening, and only assessed with smear microscopy [17], produced much lower estimates of TB prevalence (79.4 per 100,000). These differences in methods prevent direct comparisons with the 2015–2016 national prevalence survey. Second, our study is restricted to individuals 15 years and older as defined by the individual level inclusion criteria of the national prevalence survey. However, a previous study showed that the paediatric TB epidemic in Bangladesh is substantial [38]; studying spatial patterns in people under age 15 would supplement our work. Third, we only described the spatial distribution of active TB, not latent TB infection. Yet, eliminating TB will require the country to

effectively tackle the reservoir of latent infection [39]. Fourth, our analysis did not differentiate between drug-susceptible TB, multidrug-resistant (MDR) TB, and extensively resistant TB, despite distinct policy implications. Bangladesh has a high MDR-TB-burden according to the WHO, and fine-scale estimates of MDR-TB prevalence could guide priority-setting for increased access to drug sensitivity testing and second-line treatment regimens [40]. Fifth, in the absence of sub-national prevalence data for well-established TB risk factors, including smoking, alcohol use, and diabetes, we had a limited number of auxiliary variables that could be used to enhance the precision of our prevalence estimates. We estimated the prevalence of overcrowding at the district level, using a Bayesian spatial model and Bangladesh DHS 2007, 2011, and 2014 survey data. This constructed exposure was included, along with average daily temperature, in the best model according to one model selection procedure (CPO). Sixth, while our study reinforced earlier evidence of substantial gaps in case notification in many districts of Bangladesh, our interpretation lacked the tools to disentangle the causes of these gaps – treatment-seeking behaviours, accessibility of services, or reporting completeness – and their respective importance merits further investigation [41, 42]. Finally, our analysis relies on data from 2016, and the disruptions in TB services caused by COVID-19 in Bangladesh may have changed the geographical patterns described in the present work [43].

We envision several directions to build on this work. First, the results could be leveraged to improve existing methods that derive incidence from prevalence surveys by assuming a fixed relationship between prevalence, incidence, and duration of disease. Duration of disease is longer in undiagnosed cases than in treated patients. Therefore, the relationship between incidence and prevalence may be distorted in areas with large numbers of TB cases missing from care. Second, employing an integrated modelling approaches linking transmission dynamics and health systems has the potential to provide invaluable insights into the understanding of the optimal delivery of TB services [44]. Sub-national estimates of TB prevalence and case notification rates, along with local estimates of access to quality TB services from health facility survey data, could be used as inputs to represent spatial heterogeneity in these

mathematical models [45]. Third, comparing spatial patterns in active TB with sub-national variations in the distribution of the bacterial strains that cause TB could improve our understanding of the infectiousness of these strains [46]. Finally, the method presented in this study could be replicated in other high-priority settings with case notification counts and recent prevalence surveys, such as the Philippines (2016), Vietnam (2017), Myanmar and Namibia (2018), South Africa (2019), and India (2021).

Conclusion

In conclusion, our study shows that in countries, such as Bangladesh, that have achieved considerable progress in case notification and treatment coverage, substantial within-country inequalities on key indicators of the TB epidemic are still observed. Our analysis suggests that the large numbers of people who develop TB and are missing from care each year are concentrated in select districts. Enhanced routine surveillance activities along with active case-finding strategies will be key to reach, diagnosis, and treat people with TB in these areas that likely face barriers in accessing testing and treatment services. Ending TB in Bangladesh by 2035, in alignment with the WHO and UN strategies, will require closing the gaps in diagnostic, prevention, and treatment efforts. sub-national estimates of TB prevalence and prevalence-to-notification ratios can guide these efforts, help develop tailored strategies, and ensure that no population is left behind in the fight against TB.

Acknowledgements

The 2015-2016 National sero-prevalence survey depended upon the dedicated efforts of the many individuals who collected the data and who worked to assure data quality, including health care workers, disease reporting officers, regional strategic information officers, and the National Tuberculosis Control Programme implementing partners. Without their efforts, this analysis would not have been possible.

Chapter 3

WHAT DO WE KNOW ABOUT READINESS AND PROCESS QUALITY OF CARE IN LOW- TO MIDDLE-INCOME COUNTRIES? A STATISTICAL FRAMEWORK TO MEASURE SUB-NATIONAL INEQUALITIES IN READINESS AND PROCESS QUALITY METRICS USING HEALTH FACILITY SURVEYS, IN SENEGAL, KENYA, AND TANZANIA

Abstract

Background: Poor-quality care is a major barrier to health gains in most low- to middle-income countries. Attempts to measure health services quality, to date, have focused on cross-sectional, national studies, which overlooks temporal trends and sub-national inequities in healthcare quality. Yet, time series of spatially dis-aggregated estimates are necessary to assess progress towards universal access to health care of sufficient quality. Large standardized health facility surveys, such as the Service Provision Assessment (SPA) and the Service Delivery Indicators (SDI), provide the most consistent and comparable description of the level at which patients interact with the health system.

Methods: In this study, we present a statistical framework to analyze all publicly available survey data in each country, to explore spatio-temporal variations in metrics of readiness and process quality. We operationalize the World Health Organization's Service Readiness Assessment framework and the Integrated Management of Childhood Illness chart to derive metrics of readiness and process quality of sick child care from inspections of facilities and observations of sick-child consultations. We apply our framework to three countries with several rounds of SPA and SDI surveys- Senegal, Kenya, and Tanzania. Our modelling approach allows us to supplement direct survey measurements with model-based predictions of readiness and process quality metrics, which used correlations in space and time, and auxil-

iary information from spatially resolved covariates (such as health workers density or metrics of urban development) to improve the precision of our estimates for areas with sparse data.

Results: We found an overall good out-of-sample predictive validity of our models, with relatively high coverage and low mean-squared errors, across metrics and countries. Our sub-national estimates evidenced overall gaps in readiness and process quality metrics in Senegal, Kenya and Tanzania, with substantial within-country disparities.

Conclusion: Our study introduced new statistical methods that enhance the measurements of actionable quality metrics, at a sub-national resolution that can help inform programmatic decision-making and enhance equity in availability of high-quality services.

Keywords: readiness, process quality of care, healthcare quality measurement, Service Provision Assessment, Service Delivery Indicators, survey analysis, spatial analysis, health inequities

Introduction

It is estimated that more than 8 million deaths every year in low- to middle-income countries (LMICs) should be treatable by the health system [2]. Measuring healthcare access and quality is key to accountability and action; yet few LMICs have nationally representative and timely data systems to monitor access to health services of sufficient quality [3]. In their absence, health system assessments have mostly focused on metrics of geographical [47, 48, 49] and financial accessibility to health services or health inputs [50], such as the number of health facilities or beds per 1,000 inhabitants. While both sets of metrics provide useful insights into the ability to receive or provide health services, neither capture directly the quality of the healthcare received [51, 52]. Mortality-based metrics are another attempt to assess health systems performance - through excess mortality [4]. However, these measures are not suited to identify specific gaps in the cascade of care as they lump access, utilization and quality into a single metric. Therefore, improving health system quality requires new methods that enhance the measurements of actionable quality metrics [53], dis-aggregated by geographies to ensure that progress towards universal health coverage (UHC) means equitable access to quality healthcare for all [54].

While investments in routine health information systems are improving their timeliness and trustworthiness [55], health facility surveys, including the Demographic and Health Survey's Service Provision Assessment (SPA) and the World Bank's Service Delivery Indicators (SDI), remain the prime data source to assess the quality of available health services in LMICs [5, 56, 57, 58, 7, 59]. Such health facility surveys provide detailed information about various dimensions of the supply of health services in facilities, including components of structures (basic amenities, infection control, equipment, diagnostics, and medication), processes (components of clinical care), and patient perspectives (satisfaction with services received) [60]. Several recent analyses of the SPA or SDI surveys have operationalized assessment guidelines or clinical charts provided by the World Health Organization (WHO), such as the service readiness index (SRI) [61] or the Integrated Management of Childhood Illness

(IMCI) diagnostic protocols [62], to derive facility-level metrics of readiness and/or process quality of care [63, 64]. These studies have built on the strength of either the SPA or SDI to construct composite metrics of readiness and process quality comparable across countries [5, 56, 7, 59]. However, these analyses have been restricted to one type of survey (either SPA or SDI) and to the most recent year of data, and thus cannot shed light on time trends in the availability of quality care. Additionally, these studies report country-level estimates of quality. Recent work has highlighted large disparities in the coverage of key health services [65] and in health outcomes [66, 67, 68] between the administrative units of sub-Saharan African countries. These evidence of sub-national inequalities in health system output (coverage and outcomes) emphasize the need to better understand and monitor sub-national disparities in health system inputs and processes.

In contrast with previous studies, in this work we present methods to combine facility data from multiple survey years and designs. We develop Bayesian hierarchical models to analyze all publicly available health facility surveys in three countries where several rounds of SPA and SDI were conducted - Senegal, Kenya, and Tanzania - to estimate sub-national metrics of the availability of sick-child services of sufficient quality, in public and private facilities, over time and across policy-relevant units - departments, counties and regions, respectively (henceforth referred to as regions). Additionally, we present summary measures of regional inequalities in readiness and process quality metrics and show trends in each country. To our knowledge, this study is the first to describe elements of readiness and process quality of sick child care sub-nationally and over time, leveraging all available health facility survey data and using statistical methods to account for differences in survey tools. By borrowing information from previous years and adjacent geographic units [69], or by leveraging the information from auxiliary variables, a model-based approach to readiness and process quality metrics estimation allows to generate estimates at more granular resolution than direct survey estimates, and to predict estimates with associated uncertainty in administrative units where no observation was collected, and thus, no direct estimate is available. Spatio-temporal metrics of quality of care can add considerable value to the field of readiness and process

quality measurement: they can help identify geographic inequities in health service delivery; be used to measure the impact of organizational factors or interventions on the quality of health services, and provide additional evidence to the causal link between certain processes of care and specific health outcomes [70, 71]. Practically, regional-level composite metrics of readiness and process quality can be used by program managers and regional health officials to identify local bottlenecks in the supply of health services, and to target facility inspections, which could reveal gaps in the provision of certain medicines or adherence to specific guidelines [72].

Methods

Health facility survey data

Data comes from two standardized health facility assessment tools, the Service Provision Assessment and the Service Delivery Indicators. WHO's Service Availability and Readiness Assessment surveys were not included in this analysis, as they are not publicly available and did not collect information on process quality. The SPA and SDI are consistent and comparable health facility surveys nationally representative of the formal health sector. The SPA surveys analyzed in this article are based on a stratified survey design by facility type (hospital, clinics, health centers), managing authority (public/private), and first administrative division, and includes four modules: an inventory questionnaire, observations of consultations, exit interviews with the observed patients, and interviews with healthcare providers. The SDI is a stratified survey by urban/rural areas and first administrative division, and comprises three modules: an inventory questionnaire, clinical vignettes to assess providers' knowledge, and unannounced visits to facilities to measure providers' absenteeism (see supplementary methods B.1 for more details about each survey's sample and design).

In this analysis, we utilized all available cross-sectional facility surveys in Senegal, Kenya and Tanzania (see Table 3.1). We selected these three countries as several years of survey data from both SPA and SDI made it possible to estimate levels and trends in quality of

care over time and sub-national units. We focused on child health as it is the only health service for which both SPA and SDI include variables on readiness and process quality. We analyzed data from 10,384 facilities, 18,840 direct observations of sick-child consultations and 6,498 clinical vignettes, over 19 years in Kenya (1999-2018), and 10 years in Senegal (2010-2020) and Tanzania (2006-2016) (see details in supplementary tables B.1 and B.2). Facility-level information relating to the availability, readiness and quality of child health services were derived from each survey. Specifically, we extracted the 48 tracer items that form the general service Readiness Index [58], the 12 tracer items expressly included in the child-services-specific SRI, fifteen diagnostic protocols corresponding to systematic assessments and physical exams (see table 3.2) that should be performed during every sick-child visit according to the IMCI guidelines [62], and three treatment guidelines for children diagnosed with diarrhea, pneumonia or malaria . Additionally, we collected facilities' geographic location and sampling weights, and the survey design variables - region, facility type, and managing authority.

Quality of care metrics

Metrics of readiness comprised WHO general and child-services-specific SRI (henceforth referred to as general SRI and child SRI). These indices generate information on the availability of basic equipment, amenities, essential medicines, and diagnostic capacities, in the delivery of key health services, including child services, in health facilities. Both metrics of readiness were calculated for each facility as the proportion of tracer items observed and reported by the survey team. As some items were missing for SDI and older SPA surveys, we modified the two indices to only use the lowest common denominator between surveys, to ensure comparability across years in each country (see table 3.2).

We derived metrics of process quality from the content of sick child consultations observed in the SPA surveys, and the clinical knowledge displayed by health providers sampled for vignettes in the SDI surveys. In supplementary sections B.1.3 and B.1.4, we present the sampling of providers within SPA and SDI, for direct observations and vignettes respectively,

Table 3.1: List of surveys available in Senegal, Kenya and Tanzania

Senegal	Kenya	Tanzania
SDI 2010[a]	SPA 1999	SPA 2006
SPA 2012-13[b]/2014[b]	SPA 2004	SDI 2010[a]
SPA 2015[b]/2016[b]	SPA 2010	SPA 2014
SPA 2017	SDI 2012	SDI 2014
SPA 2018	SDI 2018	SDI 2016
SPA 2019		

[a]The 2010 SDI surveys in Senegal and Tanzania were pilot studies, which only sampled a small number of public facilities in selected regions of the country, and were therefore excluded from the main analysis.

[b]Dependent sampling structure between the first four rounds of the continuous SPA SPA-survey in Senegal 2013-2016 (see appendix B.2.2).

[c]SPA 2012-13[b]/2014[b] and 2015[b]/2016[b] were analyzed jointly to mitigate the effect of the dependent sampling on estimates comparability.

and we discuss the differences of assessing process quality using one or the other. Following previous studies [7, 70], adherence to IMCI diagnostic and treatment protocols were used as a proxy for process quality of sick-child care. IMCI guidelines provide the timely and appropriate diagnosis and treatment of common childhood illnesses, including pneumonia, malaria, and diarrhea [73]. Previous work has shown that IMCI is cost-effective to improve child care and reduce child mortality [74]. Additionally, Senegal and Tanzania in 1996, and Kenya in 1999 were among the first countries in the world to introduce IMCI into their national health systems [75]. Our metrics of process quality of care included an indicator of providers' adherence to 15 IMCI diagnostic protocols (as a proportion), and a measure of compliance (as a binary indicator) with recommended treatment guidelines for children diagnosed with pneumonia, malaria or diarrhea (see table 3.2), the three main killer diseases for children in sub-Saharan Africa [59].

Table 3.2: List of tracer items and diagnostic protocols used to derive readiness and process quality of care

Readiness	Process quality of care
1. General SRI: Facility has functioning	**3. IMCI diagnostic protocols**: Provider
Basic amenities: improved water source, adequate toilet	Physical examination: took child's temperature,
Power (electricity or generator), computer with internet, phone	checked for pallor, looked in child's ear, counted
Basic equipment: adult scale, child scale[a], stethoscope[a],	respiration, checked skin turgor for dehydration,
thermometer[a], light source, and blood pressure apparatus	undressed child to examine, weighed the child
Standard precautions for infection prevention:	pressed both feet to check for oedema
soap, alcohol hand rub, sharps box, latex gloves,	History taking: Provider asked for
disinfectant, disposable syringes, pedal waste bin, guidelines,	cough or difficulty breathing, diarrhea, fever,
Diagnostic capacity: for malaria (RDT kit or smear with microscope) [a]	inability to drink anything, vomiting, convulsions
Essential medicines: amoxicillin[a], oral re-hydration salt[a], zinc[a]	
ampicillin, atenolol, ceftriaxone, diazepam,	
gentamicin, insulin, magnesium, oxytocin,	
salbutamol, simvastatin, thiazide	
2. Child curative health SRI:	**4. Appropriate treatment**:
Measuring tape, infant scale,	Child diagnosed with pneumonia given antibiotic,
Co-trimoxazole, paracetamol,	Child diagnosed with diarrhea given ORS,
antimalarial drugs	Child diagnosed with malaria given treatment

[a]Items also used to calculate SARA service-specific child health readiness index

Model for sub-national estimation of quality of care metrics

The four composite indices of readiness and process quality of care described in table 3.2 were modelled separately using a small area estimation approach. Specifically, we adapted a previously developed Bayesian framework, which models the direct survey estimates as a function of covariates, and space and time components [76]. This Bayesian spatial model allows to estimate time series of the true underlying values of the quality of care metrics by smoothing over time and space the direct estimates obtained from multiple surveys, with potentially different designs and associated uncertainty. Specifically, in the first stage, we calculate the Horvitz-Thompson estimator of the quality metric in region i, year t, and survey s, by using the sampling weight $w_{k,its}$ associated to each facility k:

$$\widehat{p_{its}^{HT}} = \frac{\sum_k w_{k,its} y_{k,its}}{\sum_k w_{k,its}} \tag{3.1}$$

where $y_{k,its}$ is either a proportion (indices 1-3), or binary (index 4), in facility k, sampled in region i, over survey s in year t, and $w_{k,its}$ is the facility's sampling weight. At the second stage, we model a logit-transform of the Horvitz-Thompson estimator by using random-effects to enhance the precision of our estimates:

$$\begin{gathered} \text{logit}\left(\widehat{p_{its}^{HT}}\right) \sim N\left(\theta_{its}, \widehat{V_{its}}\right) \\ \theta_{its} = X_{it}\boldsymbol{\beta} + \gamma_t + \alpha_t + e_i + S_i + \delta_{it} + \nu_s \end{gathered} \tag{3.2}$$

where $\widehat{V_{its}}$ is the design-based variance of logit $\left(\widehat{p_{its}^{HT}}\right)$, X_{it} are regional-level covariates observed in region i in year t (see table 3.3), α_t are temporal random effects modeled as a first order random walk, S_i are spatially structured random effects [77] (see B.2.1), δ is a space-time interaction, and γ_t, e_i and ν_s are respectively temporal, spatial and survey zero-mean independent random effects. For each of the four indices of quality, we considered seven possible models (see table B.3) and used rigorous selection techniques to identify the best performing model (see B.2.3). Regional-level estimates and their associated uncertainty were then obtained by drawing 1,000 posterior samples for all parameters and calculating the mean, and the 2.5$^{\text{th}}$ and 97.5$^{\text{th}}$ quantiles. We assessed temporal trends in sub-national

Table 3.3: List of covariates considered in fitting equation (3.2)

Covariate	Spatial resolution	Temporal resolution	Source
Urbanicity	Pixel-level	Annual	ESA and Land Cover
Health worker density	Second-level administrative unit	Annual	Censuses (derived)
Travel time to nearest city	Pixel-level	Annual	MAP [78]
Travel time to nearest health facility	Pixel-level	Annual	MAP [49]
Total population under five years old	Pixel-level	Annual	WorldPop
Average years of education in adult population	Second-level administrative unit	Annual	IHME [79]
Night-time lights	Pixel-level	Annual	VIIRS and DMSP (harmonized)
Human development index	Pixel-level	Annual	[80]
Elevation	Pixel-level	Annual	ArcGIS

disparities using three summary measures of inequalities - the inequality pattern index, the high-to-low ratio and high-to-low difference (see section B.2.4).

To test the predictive validity of the selected models, we followed Mercer and colleagues [76]: we removed all the observations for region i in year t, and compared the predicted quality metrics and their 95% credible intervals obtained using model (3.2), with the direct survey estimate $\widehat{p_{its}^{HT}}$. We examined the coverage of our estimates, i.e. the frequency at which the direct survey estimate was contained within the model estimate 95% interval.

Results

Modelling metrics of readiness and process quality of care

Table 3.4 shows the final model chosen through our selection procedure, for each readiness and process quality metric, in each of the three countries. A first important feature of our modeling analyses was the hypothesis that spatio-temporal smoothing would improve overall model fit. We see that indeed models including spatially-correlated random effects were preferred to model with independent random effects in all but three instances. The diagnostic protocols metric was best modelled without spatial smoothing in both Senegal and Kenya. For the general SRI metric in Kenya, a non-spatial model was also preferred. A

Table 3.4: Selected model for each readiness and process quality metric, and variance decomposition. Columns $\sigma^2_{S_i}$, $\sigma^2_{\gamma t}$, $\sigma^2_{\alpha t}$, $\sigma^2_{\delta_{it}}$, $\sigma^2_{\nu_s}$, and $X_{it}\boldsymbol{\beta}$ indicate the percentage of total variance explained by each component.

	Senegal						
Outcome	Model selected	$\sigma^2_{S_i}$	$\sigma^2_{\gamma t}$	$\sigma^2_{\alpha t}$	$\sigma^2_{\delta_{it}}$	$\sigma^2_{\nu_s}$	$X_{it}\boldsymbol{\beta}$
General SRI	5	42.1	4.0	2.6	51.2	-	-
Child SRI	6	35.4	1.4	1.4	60.4	1.4	-
Diagnostic protocols	1	5.5	2.4	2.7	89.4	-	-
Adequate treatment	7	.2	3.5	2.5	2.7	58.8	32.3
	Kenya						
Outcome	Model selected	$\sigma^2_{S_i}$	$\sigma^2_{\gamma t}$	$\sigma^2_{\alpha t}$	$\sigma^2_{\delta_{it}}$	$\sigma^2_{\nu_s}$	$X_{it}\boldsymbol{\beta}$
General SRI	1	2.1	53.8	42.7	1.3	-	-
Child SRI	5	1.6	69.2	26.4	2.7	-	-
Diagnostic protocols	3	3.1	1.3	2.2	64.2	-	29.3
Adequate treatment	5	2.8	96.9	.1	.2	-	-
	Tanzania						
Outcome	Model selected	$\sigma^2_{S_i}$	$\sigma^2_{\gamma t}$	$\sigma^2_{\alpha t}$	$\sigma^2_{\delta_{it}}$	$\sigma^2_{\nu_s}$	$X_{it}\boldsymbol{\beta}$
General SRI	5	0	6.4	84.2	9.4	-	-
Child SRI	7	0	6.3	5.2	77.0	6.3	5.0
Diagnostic protocols	7	.1	4.3	3.9	57.2	9.6	24.8
Adequate treatment	6	0.6	21.5	0.4	34.8	42.7	-

second important feature of spatio-temporal modeling resides in the strengths of the different sources of variability. Table 3.4 reveals important differences in the contribution of each component of the model, between indicators and countries. Spatial and and spatio-temporal variability accounted for more than 90% of the variability in the general SRI, child SRI and diagnostic protocols metrics in Senegal, for more than two-third of the variability in the diagnostic protocols and one-third of the adequate treatment metrics in Kenya, and for 77.0% and 57.2% of the variability in the child SRI and diagnostic protocols metrics in Tanzania. Spatial variability suggests a strong influence of geographic factors, while spatio-temporal variability, which amounts to the interaction between space and time, indicates changes in the spatial distribution of the metric, with high areas becoming low, and conversely [81]. Strong spatio-temporal variability is typical of the gradual roll-up of interventions or programs, affecting different regions over the study period. Strong temporal variability, on the other hand, is indicative of little geographic variability but with a strong temporal trend that changes the metric across sub-national units uniformly over the study period. Temporal variability explained over 80% of the variability in the general SRI metric in Tanzania, where all sub-national units experienced large gains in general readiness, over the study period (see supplementary figure B.6). It also accounted for a significant share of the variability in the general SRI and child SRI metrics in Kenya, where an overall positive trend was estimated over the study period (B.5), but with significant volatility from one year to the next, as illustrated with the high contribution of the zero-mean temporal random effects in explaining the total variance. Finally, a substantial proportion of the variation in the adequate treatment metric in Senegal (58.8%) and Tanzania (42.7%), was explained by the survey random effects - accounting for differences in sampling or survey instrument, respectively. In Senegal, this could be explained by the change in the sampling frame for private facilities for 2017, which led to a substantial drop in the national estimate of adequate treatment in private facilities, while in Tanzania, this results likely reflects the important difference between assessing providers' practices from vignettes rather than direct clinical observation (see section B.1.4 for more details). Regional-level covariates were selected in four of the twelve models - three

of them fitting process quality metrics, and explained a substantial proportion of the variance in these metrics. The fact that only four out of the twelve models included regional-level covariates highlights the general challenge in demonstrating associations at the ecological level, and the specific challenges of finding strong predictors of healthcare quality. To assess

Table 3.5: Performances of the models, assessed using mean squared errors (MSE) and spatial leave-one-out predictions for the four metrics, across Senegal, Kenya and Tanzania

	Senegal			
Criteria	General SRI	Child SRI	Diagnostic Protocols	Adequate treatment
MSE	4.0×10^{-3}	5.5×10^{-3}	6.0×10^{-3}	3.6×10^{-2}
Coverage (%)	83	79	88	77
	Kenya			
Criteria	General SRI	Child SRI	Diagnostic Protocols	Adequate treatment
MSE	2.9×10^{-3}	3.7×10^{-3}	1.2×10^{-2}	2.6×10^{-2}
Coverage (%)	88	86	85	83
	Tanzania			
Criteria	General SRI	Child SRI	Diagnostic Protocols	Adequate treatment
MSE	2.5×10^{-3}	4.3×10^{-3}	5.4×10^{-3}	2.9×10^{-2}
Coverage (%)	80	92	92	94

the accuracy of our models' predictions, we used the mean squared error (MSE), and the coverage for the left-out areas, as the proportion of the direct survey estimates contained in the model-based 95% credible intervals. The results in 3.5 show that the MSE is higher for the process quality metrics than the readiness metrics, across the three countries. Addition-

ally, although coverage is generally high, above 80% for all but two models, it seems slightly worse in Senegal, especially for the child SRI and adequate treatment metrics. However, supplementary figures B.1-B.2-B.3 suggest that these marginally better coverage indicators for Kenya and Tanzania come at the expense of larger 95% credible intervals, especially in Tanzania.

Levels and trends in readiness and process quality metrics

Figure 3.1 is a four-quadrant graph comparing sub-national estimates of general and child SRI across sub-national untis, for each country, in 2010 and 2020. In Senegal, we see that if estimated readiness has improved between 2010 and 2020, with improvements in general SRI across sub-national units, only one region was in the upper-right quadrant of figure 3.1, i.e. above 80% on both readiness metrics in 2020. Nevertheless, about half of the sub-national units achieved a child SRI above 80% in 2020, and spatial disparities seem to have decreased compared to 2010. Kenyan sub-national units have shifted to the left on figure 3.1 between 2010 and 2020, suggesting a decrease in general SRI across sub-national units, while the spread of the cloud of points has increased, which indicates larger sub-national disparities in 2020. In Tanzania, readiness is estimated to have improved as a result of gains in child SRI, and in general SRI, to a lesser extent, across sub-national units. Figure 3.2 presents maps of the mean and length of 95% interval estimates of adherence to IMCI diagnostic protocols, in Senegal, Kenya and Tanzania in 2020. This process quality metric displayed substantial spatial variability; average compliance to IMCI protocols ranged from 26.2% (14.7-42.2) in Gossas to 61.6% (43.3-77.2) in Guinguineo in Senegal, from 32.9& (18.1-52.2) in Vihiga to 72.0% (53.4-85.3) in Garissa in Kenya, and from 23.8% (8.3-51.9) in Morogoro to 45.6% (20.0-74.2) in Dodoma in Tanzania. Uncertainty is much greater in Tanzania than in Senegal, and to a lesser extent, than in Kenya, as highlighted by the width of the credible intervals on figure 3.2**b-d-f**, which can be explained by the recency of the last data collection - 2016 in Tanzania, compared to 2018 in Kenya and 2019 in Senegal. We also examined the joint distribution of adherence to diagnostic protocols and adequate treatment to examine spatial

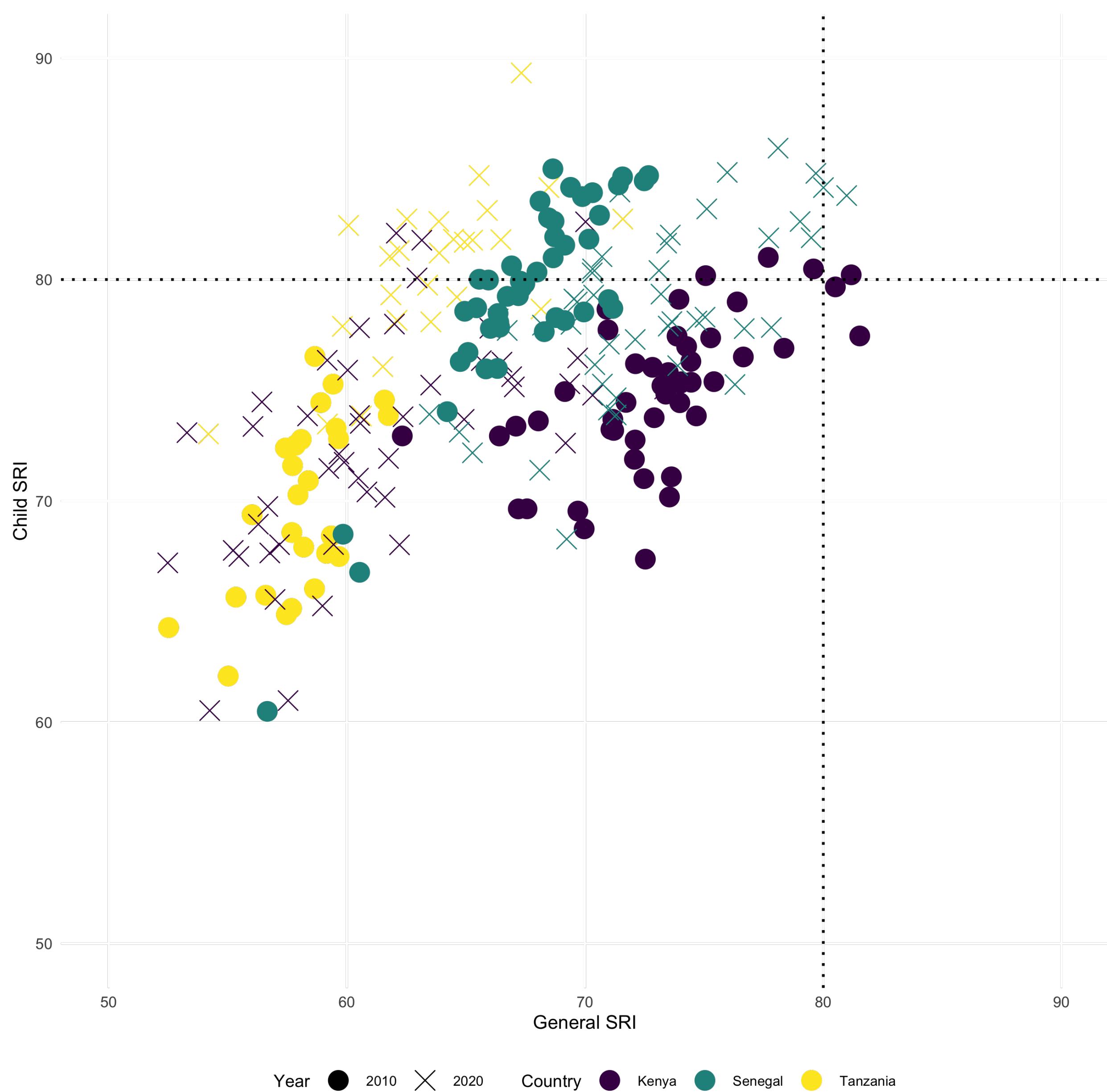


Figure 3.1: Estimated readiness metrics (general Service Readiness Index and child Service Readiness Index) across sub-national units in Senegal, Kenya and Tanzania, in 2010 and 2020

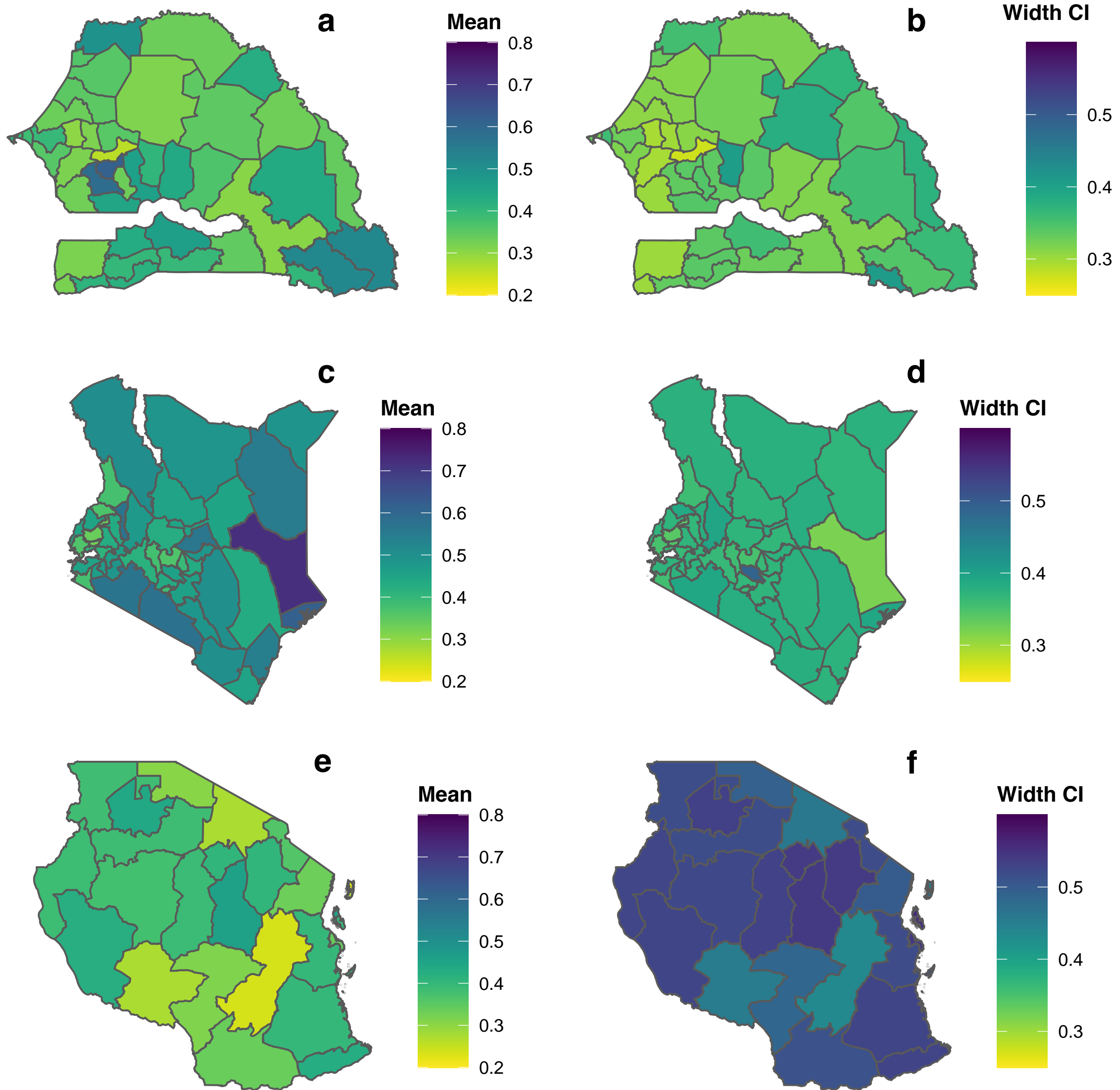


Figure 3.2: Mean estimates of adherence to IMCI protocols (a,c,e), and length of the 95% interval (b,d,f) estimates, in Senegal (a,b), Kenya (c,d), and Tanzania (e,f) in 2020

correlations in these two metrics (see supplementary figure B.7). Bi-variate maps grouping sub-national units by metrics' quantiles were used to describe regions' relative performances

over the study period, and identify priority areas. We noted an increase in the proportion of regions ranking either in the top or bottom quantile for both dimensions of IMCI diagnostic and treatment protocols, over the study period. This may suggest reinforcing dynamics creating a larger gaps between on the one hand, sub-national units where, on average, sick children are more likely to be diagnosed according to appropriate standards and received an adequate treatment, and on the other hand, regions where both providers' diagnosis and treatment practices tend to differ from IMCI protocols.

We assessed trends in disparities in each readiness and process quality metrics, separately, by estimating three summary measures of inequalities (see table 3.6) - the inequality pattern index, the high-to-low ratio and the high-to-low difference, for the three countries, in 2010 and 2020. Across inequalities summary measures, process quality metrics were likely to be more unequally distributed between sub-national units than readiness metrics in 2010 and 2020. In Senegal, however, the high-to-low difference was largest for the general SRI. This is consistent with the results displayed in figure 3.1, where large absolute differences in general SRI were evidence for Senegal, while in Kenya and Tanzania the clouds of sub-national units were more concentrated. Inequalities in readiness metrics tended to increase over the study period, in all three countries. In Senegal, inequalities in process quality metrics decreased marginally, while in Tanzania they increased substantially. In Kenya, the two process quality metrics had opposite inequality trajectories: the difference in adherence to IMCI protocols between the worst and best performing sub-national units grew by 13.7% between 2010 and 2020. Conversely, the difference in adequate treatment decreased by 21.0%, over the same period.

Gaps in composite metrics

Metrics at the regional-level can mask important differences in the readiness and process quality across facilities within a region. For instance, a regional average metric of 80% may reflect homogeneous facilities with every score around 80%, or a bi-modal distribution of 60% and 100% scores in facilities. To observe potential differences in the underlying distribution of

Table 3.6: Summary inequality measures, across sub-national units of Senegal, Kenya and Tanzania

Measure	General SRI		Child SRI		Diagnostic Protocols		Adequate treatment	
	Senegal							
	2010	2020	2010	2020	2010	2020	2010	2020
Inequality pattern index (%)	−4.1	−13.5	-7.3	-5.0	3.1	7.9	-62.9	-23.5
High-to-low ratio	1.7	1.9	1.7	1.6	3.4	3.1	4.5	2.4
High-to-low difference (%)	31.9	39.0	8.2	6.1	12.7	11.1	30.8	12.4
	Kenya							
	General SRI		Child SRI		Diagnostic Protocols		Adequate treatment	
	2010	2020	2010	2020	2010	2020	2010	2020
Inequality pattern index (%)	-1.6	3.8	−2.3	.1	1.8	10.1	11.6	-.1
High-to-low ratio	1.3	1.5	1.3	1.5	2.2	3.0	7.5	2.4
High-to-low difference (%)	20.4	26.0	18.4	26.5	35.2	48.9	68.1	47.1
	Tanzania							
	General SRI		Child SRI		Diagnostic Protocols		Adequate treatment	
	2010	2020	2010	2020	2010	2020	2010	2020
Inequality pattern index (%)	−4.6	−4.6	.1	−14.5	-2.8	-4.8	-15.3	-5.9
High-to-low ratio	1.5	1.7	1.7	1.8	2.3	3.6	4.2	20.0
High-to-low difference (%)	20.7	31.1	35.0	40.4	29.5	43.1	53.8	64.2

facilities' performances, we present the proportion of facilities with a general SRI below 50%, between 50% and 80%, and above 80% on figure 3.3. Interestingly, figure 3.3 **a-b** suggests that the improvements in the average general SRI across sub-national units (evidenced with figure 3.1) has resulted from an increase in the proportion of facilities with scores above 80%. While these high-performing facilities represented a minority of facilities in all but one region in 2013, in 2019, they represent a majority of facilities in fourteen regions, including three regions where more than 80% of the facilities scored above 80%. The color key used for each year is displayed on the right-hand side of figure 3.3 **a-b**; each dot represents an individual facility. With the contour plots drawn on the color keys, we see that between 2013 and 2019 the distribution of facilities' general SRI has shifted from most facilities with scores between 50-80% to most facilities with scores above 80%.

Furthermore, health facility surveys are designed to be geographically representative, but temporal variations are not accounted for in the survey design. Therefore, we investigated the possibility that seasonality could influence the availability of certain items or compliance with certain diagnostic protocols. In particular, given that Senegal and Tanzania are countries with high malaria prevalence, with distinct transmission patterns, we thought of testing the hypothesis of seasonality in providers' fever history assessment- a proxy for malaria. We divided each country into two epidemiological zones - an "endemic", and a "seasonal". In Senegal, the tropical zone in the south and southeast (regions of Ziguinchor, Kolda, Kedougou, Tambacounda and Sedhiou) with year-round transmission peaking during the rainy season (from June to October) and lower transmission during the rest of the year was labelled as "endemic", while the Sahelian zone in the north, with higher transmission toward the end of the rainy season (from July to September) and very low transmission during the rest of the year was labelled as "seasonal". In Tanzania, this division is artificial as year-round transmission affects 80% of the countries' regions; we divided the country into two zones based on malaria prevalence. Our analyses consisted of comparing the weighted frequency of fever history assessment, by providers during sick child consultations, by epidemiological zone and season in Senegal and Tanzania. Our hypothesis was that evidence of seasonality

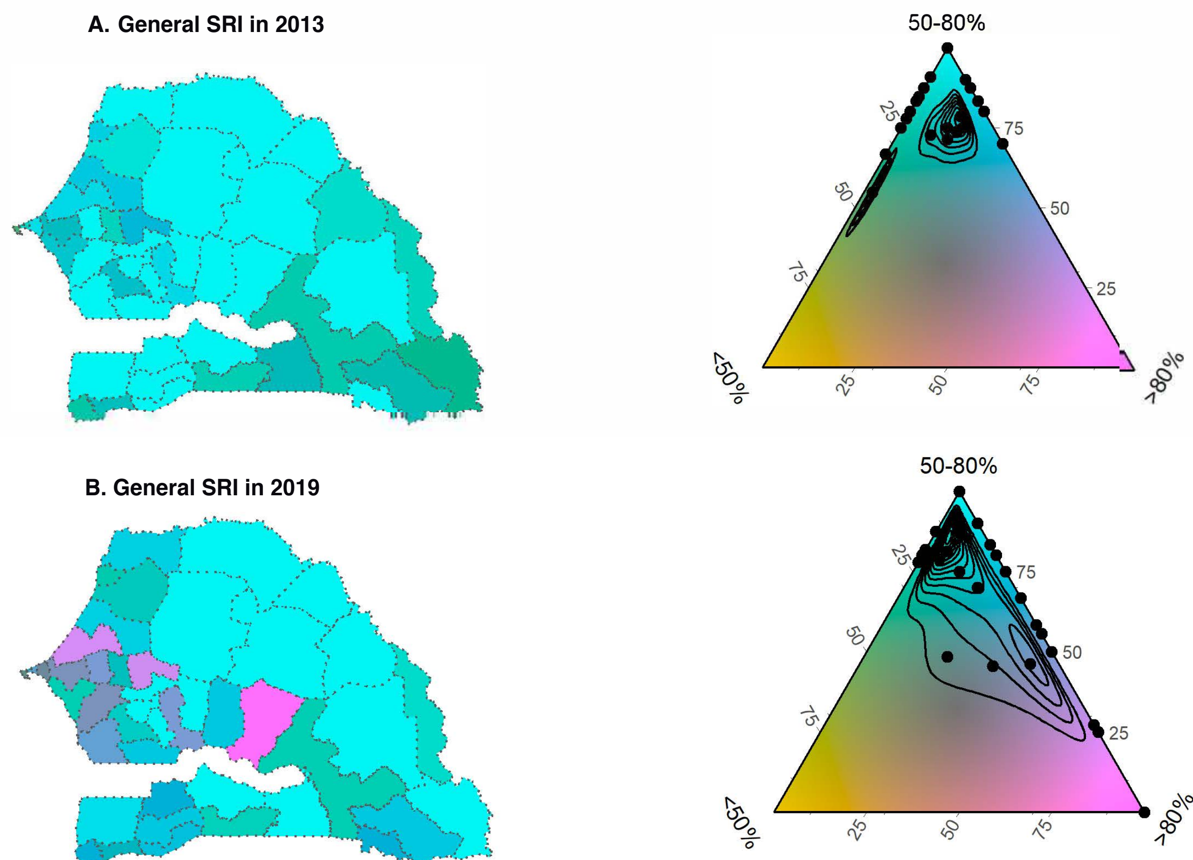


Figure 3.3: Facilities' general SRI in Senegalese departments, in 2013 (a) and 2019 (b)

would mean higher frequencies of fever assessment in the rainy season compared to the dry season, especially in the regions where malaria is seasonal. As malaria is endemic in most of Tanzania, we expected to see stronger evidence of seasonality in Senegal than Tanzania. We find that in most years, assessment of fever history is more frequent during the rainy season, especially in facilities located in the "seasonal" epidemiological zone (see figure 3.4). However, 2018 is one notable exception where the assessment was done more frequently during the "dry" season, in both zones. This discrepancy with other years may be related to the fact that the 2018 SPA survey occurred during the health worker strike in Senegal, which may

have affected different regions differently. Figure 3.5 shows a similar pattern in Tanzania, where fever history assessment was significantly more frequent during the rainy season in 2014, in both zones, and in 2006 in the endemic zone.

Discussion

In this study, we expanded on previous healthcare quality measurement works in several important ways. First, while previous studies had focused on cross-sectional country-level analyses of health facility surveys, we presented statistical methods to jointly analyze all available facility assessments, including several rounds of surveys from the same or from different data collection instruments, which allows to explore temporal trends and spatial disparities in readiness and process quality metrics. We applied our modelling approach to all publicly available data from two main multi-country facility survey instruments, the SPA and SDI, in three LMICs - Senegal, Kenya and Tanzania. Our study demonstrates the benefits of using a modeling approach alongside direct survey measurements to estimate metrics of readiness and process quality. Spatio-temporal smoothing leverages correlation structures in space and time to improve the precision of estimates when data are sparse [82], while covariates further improve precision by separating the sources of variability underlying the temporal and geographic distribution of our indicator of interest [81]. Therefore, the borrowing of information across areas and years, and the use of auxiliary variables, allowed us to provide estimates at a policy-relevant resolution – departments, counties and regions – a spatial scale for which most surveys were not powered to report reliable estimates. Additionally, in years or regions where no data was collected, our approach 'filled-in' data gaps, by providing estimates of readiness and quality metrics with associated uncertainty. This feature of our approach is particularly appealing, as it provides means to supplement direct estimates derived from surveys that are typically conducted irregularly, with readily available and continuously collected spatially and temporally indexed covariates. Our spatial leave-one-out procedure, designed to test the accuracy of our models' predictions, suggested a high predictive validity, with coverage ranging from 77% to 94%. Variance decomposition

provided further insights into the various sources of variability in each readiness and process quality metrics. For instance, strong spatio-temporal variability in Senegal suggested important variations in the distribution of low and high readiness and process quality metrics between departments, over the study period, which may reflect the effects of geographically targeted interventions during that time, such as the USAID priority actions in certain regions of concentrations [83]. Moreover, we noted that the survey random effects, indicating differences in survey instruments in Tanzania, and in the sampling frame in Senegal, explained a substantial proportion of the variations in the adequate treatment metric. In Senegal, the higher coverage of adequate treatment in the repeated sampling phase, may reflect a form of Hawthorne effect where, in facilities assessed for the second time, interviewees had a better understanding of interviewers' expectations and adapted their practices accordingly. In Tanzania, higher average performances on this metric in the SDI than in the SPA surveys likely reflect the important difference between assessing providers' practices from vignettes rather than direct clinical observation and the gap between "knowing and doing" (see section B.1.4 for more details).

Second, generating model-based estimates of readiness and process quality metrics over time and space revealed substantial and persistent gaps. To begin with, while readiness has improved in most of Senegal and Tanzania, many sub-national units were still missing a fourth to a third of the minimal inputs necessary to provide child health services in 2020. We noted large gaps in the availability of essential medicines, including antibiotics such as amoxicillin to treat pneumonia or cotrimoxazole. In Kenya, we found evidence of a significant downward slopping trend in average readiness, driven by a substantial drop in the general SRI metrics across sub-national units; based solely on SDI surveys, the 2018 report in Kenya documented evidence of a decrease in facility average readiness; our findings confirm and expand this result, using additional data. Furthermore, despite IMCI guidelines being a central component of Senegal's, Kenya's and Tanzania's strategy to reduce child mortality, we estimated low metrics of process quality. Estimated average compliance to IMCI diagnostic protocols guidelines in 2020 exceeded 50% in just a handful of sub-national units across

Senegal, Kenya and Tanzania. The assessments less commonly done during consultations were asking parents whether their child had convulsed or been unable to drink, two danger signs indicating the need to urgently refer a child to an hospital. Recent studies conducted in sub-Saharan Africa similarly reported low levels of compliance with IMCI guidelines: using cross-sectional data collected from health facilities in the Republic of Congo, Cameroon and the Central African Republic, Perales and colleagues found that providers had completed 12 of the 15 IMCI diagnostic protocols in only 8% of sick-child visits [70], while a recent analysis of the 2014 Ethiopian SPA found that the three main symptoms of child diseases were assessed in 51% of consultations, and the three general danger signs in 4% of visits [64]. Finally, we found evidence of substantial and persistent disparities in readiness and process quality between sub-national units. The high-to-low absolute difference suggested that within-country absolute differences between regions in general and child SRI could be as high as 40% in Senegal and Tanzania in 2020, which points to systematic differences in the availability of minimal inputs of health services. Process quality metrics tended to display the largest gaps between the best and worst performing areas. Absolute difference in compliance with diagnostic protocols was equal to or greater than 30% in Kenya and Tanzania. Additionally, the average likelihood of receiving adequate treatment for a child diagnosed with pneumonia, malaria or diarrhea in a given region varied up to 7.5 fold. Evidence of top inequalities combined with low national coverage for the diagnostic protocols metric in Senegal and Kenya suggest that a national effort to improve compliance with IMCI diagnostic protocols would be appropriate. In the case of general SRI and adequate treatment metrics in Senegal, where we see significant bottom inequalities, suggesting a few areas lagging far below the national mean, spatially-targeted interventions could be better-suited to resolve this pattern of inequality [84].

Third, our modelling approach can be used to critically assess the production of readiness and process quality indicators. To begin with, there is a critical need to assess the optimal frequency and scope of health facility assessments, which so far have been conducted as occasional surveys (Kenya and Tanzania), one-time census (Haiti or Malawi), or

continuous yearly survey (Senegal since 2012). Our modelling approach allows to attune data collection efforts to specific target metrics. Estimates' precision comes from adequately appraising variability: strong spatial variability and little temporal variability of a quality metric would suggest a less frequent but more geographically diverse sample, while frequent but smaller samples would be more appropriate for metrics displaying substantial temporal but low spatial variability. Looking at the average width of the relationship between data on spatially-targeted interventions or funding and data on readiness or process quality credible intervals can provide additional evidence as to where more data need to be collected to improve the reliability of our estimates. For instance, we found that, on average over the study period, the uncertainty in the readiness and process quality indicators was much larger in Kenya and Tanzania, where data are not collected as regularly as in Senegal. Furthermore, following previous studies, readiness and process quality metrics were constructed as composite indicators, a popular approach in the measurement of healthcare quality. We used established assessment frameworks, such as WHO's SARA and IMCI, to only include items relevant to readiness and process quality. However, the summary quality value provided by composite indicators can be flawed if any of its constituent parts is biased [85]. In particular, we explored the possibility that, because facility surveys assessments are designed to be representative geographically not temporally, cross-sectional analyses could miss patterns of seasonality in items' availability or protocols performed [86]. We found evidence of seasonality in providers' adherence to some IMCI guidelines, with on average a higher frequency of fever history assessment during sick-child visits in the rainy season than in the dry season in Senegal and Tanzania. Seasonality is an important threat to the comparability of readiness and process quality metrics over time; changing the timing of the facility survey from year to the next could introduce variability clouding the overall trends in the quality metrics. Another important implications of this results is that it suggests that skipping certain diagnostic protocols could reflect providers' strategic allocation of their limited consultation time, rather than a lack of knowledge of these protocols [87]. We also found that providers in regions where malaria is endemic were more likely to ask if a child had fever, reinforcing

this hypothesis. Given the potential of IMCI diagnosis protocols to reduce mortality in children aged five and less, identifying the gaps in adherence to these standards of care remain important, but understanding the reasons for non-compliance will be key to design effective recommendations to close these gaps.

Our study has several limitations. First, the SPA and SDI exclude non-formal healthcare providers, which can represent a significant component of the health system in many LMICs [88]. In Senegal, we also excluded health huts, as the health community agents that work in these single-room facilities are not trained to treat sick-children past basic care [89]. However, health huts are often in the front-line in the management of sick children in remote rural regions, and studying their referral practices to larger health facilities would be crucial [90]. Second, our results highlighted the substantial variability in the process quality metrics introduced by differences in the sampling or the survey to collect the data. Enhanced harmonization between data collection instruments, as entailed by WHO's new Harmonized Health Facility Assessment initiative, will be necessary to ensure the comparability of survey estimates. Third, the SARA and IMCI guidelines correspond to basic equipment and diagnostic protocols that should be available in every facility, and during every sick-child visit, respectively. Therefore, our metrics reflect what could be considered the provision of a minimum level of care in health facilities as recommended by global guidelines, rather than of high-quality care [7].Fourth, our composite metrics were calculated as simple sum of items or protocols, which assumes their equal importance to quality, and ultimately to health outcomes. In practice, this may not be the case and items could be assigned weights based on their contribution to health improvements. However, such an approach would require equally large assumptions about the relative importance of each item, and we opted for equal weighting, the most common approach in the literature to date [7]. A final limitation of this study is that it focuses on the readiness and process quality of health services, not their accessibility. The most recent SPA and SDI surveys provide facilities' GPS coordinates and record fees associated with the provision of care; thus, we envision that future studies could extend our approach to incorporate information about geographic and financial accessibility of care,

alongside quality. Other directions to build on our work include looking at the relationship between data on spatially-targeted interventions or funding and data on readiness or process quality, to assess the impact of macro programs or decisions on local health systems [91]. Additionally, our work could be used to investigate the role of healthcare quality variations in explaining the large sub-national inequities in under-five mortality highlighted in recent works [9]. Ideally, surveys collecting simultaneously information about facilities and the households attending them, such as GAVI Full Country Evaluations, would provide insights into barriers to access care, healthcare-seeking behaviors, and ultimately healthcare impacts on health outcomes [92]. In practice, major international standardized facility surveys like SPA or SDI are rarely coordinated with large household surveys, which leads to temporal or spatial misalignment limiting the possibility of linkage [93]. Despite these data gaps, two recent studies have attempted to estimate the effect of healthcare quality on individual child death [70, 94]. Using cross-sectional data from health facilities in the Republic of Congo, Cameroon and the Central African Republic, Perales and colleagues found lower odds of mortality for children living in districts with higher process quality, measured as compliance with IMCI guidelines during sick-child visits. Linking births recorded in the DHS household surveys with the most recent SPA in seven countries and over 20 years, Simmons et al. used the proportion of IMCI-trained staff and the proportion of facilities with a doctor by first-level administrative division to proxy process quality; they found a non-significant effect of the former variable, and a negative association between prevalence of doctors and child survival. Future research efforts could utilize our approach to readiness and process quality measurements to test these results on a larger temporal and geographic scale than Perales and colleagues, and using alternative metrics of quality, compared to Simmons and colleagues.

Conclusion

Strengthening health services will be key to reduce the large burden of deaths amenable to healthcare in low- to middle income countries. Standardized health facility assessments

are a precious resource. We present a statistical framework that allows to jointly analyze multiple rounds of surveys from different instruments, while accounting for their specific design. Building on spatio-temporal correlations and auxiliary information, our modelling approach allows us to explore temporal trends and sub-national disparities in the readiness and process quality of care metrics. As the World Health Organization has been promoting a new Harmonized Health Facility Assessment, to ensure the alignment of health facility survey instruments and enable comparability of results over time and across countries, our statistical framework fills a methodological gap to assess trends and geographic inequalities in availability, quality and effectiveness of health services, and track countries' progress toward universal health coverage.

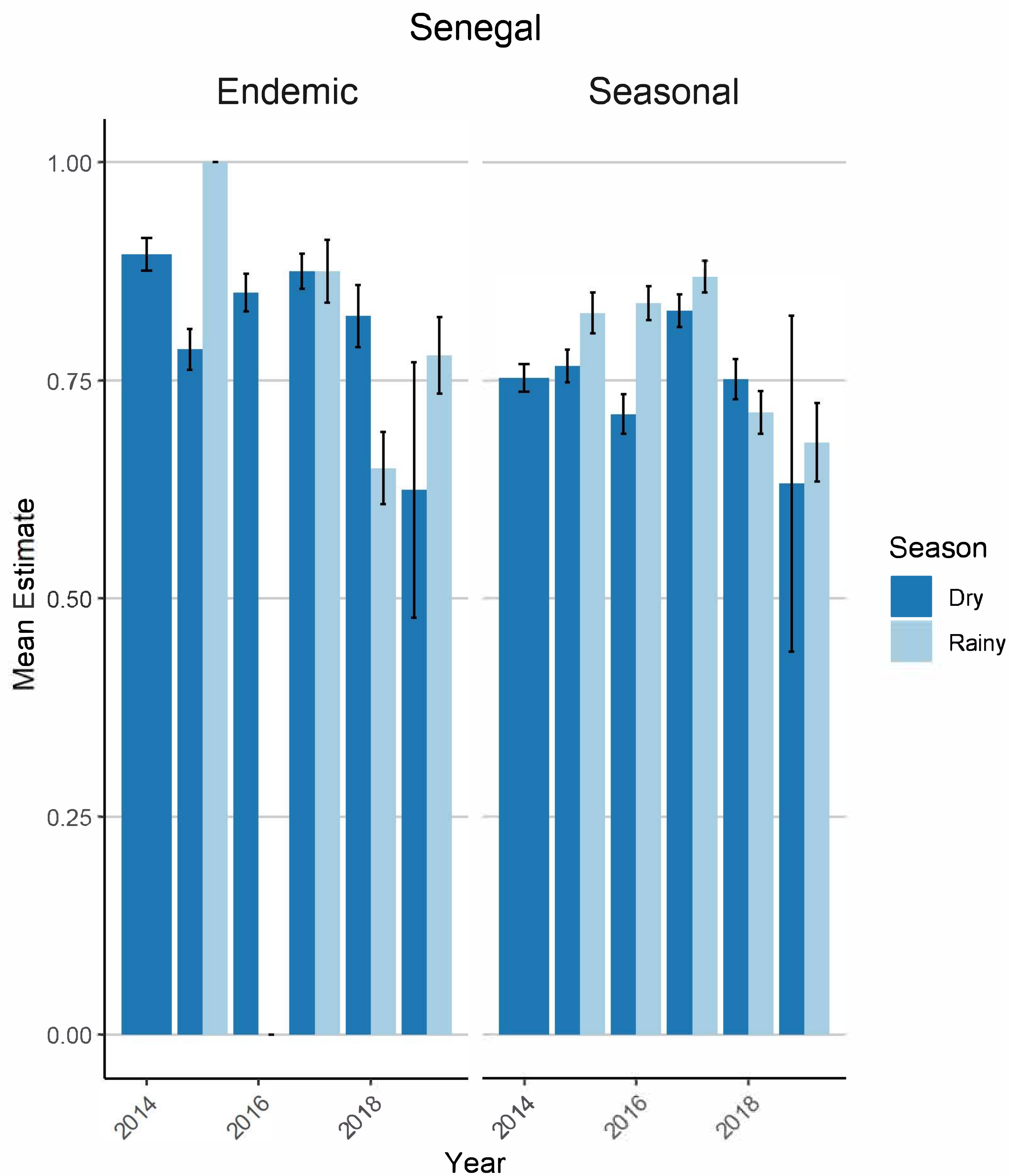


Figure 3.4: Prevalence of fever history assessment during sick-child visits in Senegal, in the departments where malaria is endemic (left panel) or seasonal (right), during the dry or rainy seasons

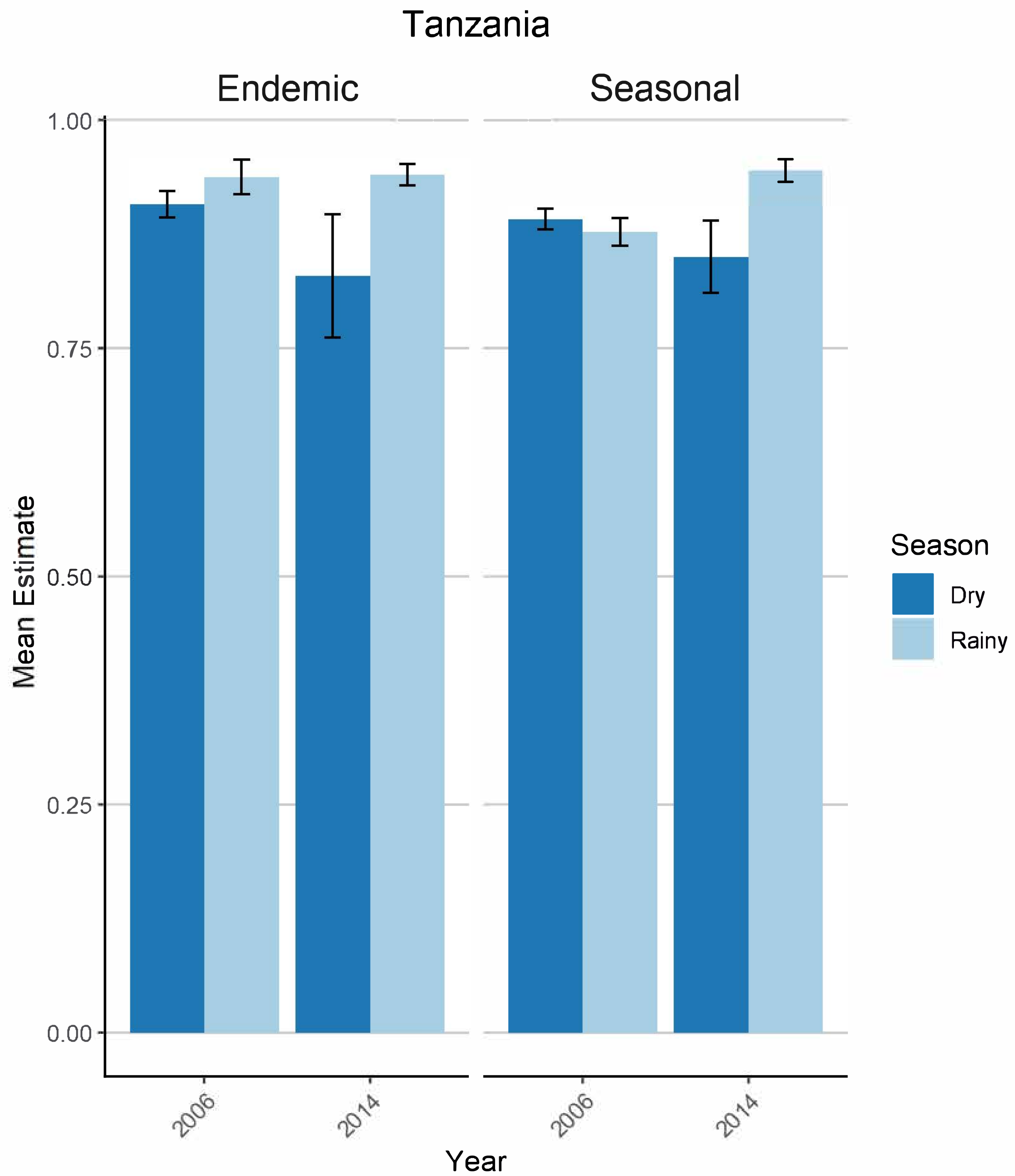


Figure 3.5: Prevalence of fever history assessment during sick-child visits in Tanzania, in the regions where malaria is endemic (left panel) or seasonal (right), during the dry or rainy seasons

Chapter 4

METRICS OF QUALITY-ADJUSTED HEALTH-SERVICE COVERAGE FOR SICK CHILDREN AND CHILD MORTALITY IN SENEGAL, KENYA AND TANZANIA: AN ECOLOGICAL REGRESSION STUDY

Abstract

Background Achieving Universal Health Coverage has the potential to significantly reduce excess mortality for conditions targeted in the Sustainable Development Goals, including child mortality, as the majority of under-5 deaths are due to causes amenable to healthcare. Quality-adjusted coverage, defined as the proportion of the population in need of a health service that receives appropriate treatment according to standards, has been described as a useful metric to track progress toward UHC.

Methods We describe temporal trends and sub-national disparities in quality adjusted coverage for sick-child care in three low- to middle-income countries; Senegal, Kenya and Tanzania. We estimate four different metrics of quality-adjusted coverage, corresponding to more or less concise approaches to estimate readiness and process quality of sick-child care, from households and health facility data, and compare their performance in predicting under-five mortality rates, over time, and at sub-national levels, using an ecological regression framework. Our model includes regional-level estimates of risk factors (such as diarrhea and lower-respiratory infection incidence) and preventive interventions (such as immunization coverage) to account for differences between regions.

Results: We found uniformly low-levels of quality-adjusted coverage across the sub-national units of the three countries, with substantial spatial disparities along the cascade of service provision (service contact coverage, receipt of intervention coverage). We reported a neg-

ative association between quality-adjusted coverage and under-five mortality rates at the ecological-level.

Conclusion: We found no evidence of one quality-adjusted coverage metric out-performing the others, which suggests that more expansive measures of readiness and process quality of care may include partially redundant information on health services.

Keywords: effective coverage; healthcare quality; survey analysis; spatial analysis; UHC

Introduction

Universal Health Coverage (UHC), defined as a universal access to health care of sufficient quality, has been described as a central piece of the Sustainable Development Goal (SDG) 3 aiming “to ensure healthy lives and promote wellbeing for all, at all ages". Achieving UHC has the potential to significantly reduce excess mortality for conditions targeted in the SDGs, including child mortality. The majority of under-5 deaths are due to causes amenable to healthcare, and a third are due to pneumonia, diarrhea, and malaria in sub-Saharan Africa [95]. Monitoring countries' progress toward UHC is however challenging. Indicators of health-service contact coverage, commonly measured for Reproductive Maternal Neonatal and Child Health (RMNCH) services in large standardized household surveys, such as the Demographic and Health Survey (DHS), have been used to assess progress toward UHC [96, 97] but lack information on quality of care, which can lead to overstating the population health benefits of these services, because of poor-quality care [2]. Effective coverage, a measure of need, use, and quality of health services, has been described as a useful metric to track progress toward UHC [51]. The lack of methodological consensus around the measurements of its three components has however limited its use in practice. Following Tanahashi's seminal work on health-service coverage measurements [98], two recent studies have laid out a framework to measure effective coverage through the measurements of a “cascade of six stages". Service contact coverage (1) is defined as the proportion of those in need of a health service who visits a health provider; input-adjusted coverage (2) represents the proportion of those in need of a health service who visits a health provider ready to provide this service; intervention coverage (3) is the proportion of those in need of a health service who comes into contact with a service that is ready and that receives the service, quality-adjusted coverage (4) is the proportion of those in need of a health service who comes into contact with a service that is ready and that receives the service according to evidence-based processes of care; user-adherence-adjusted coverage (5) is the proportion of those in need of a health service who receives the service according to evidence-based processes of

care and that adheres to provider instructions; and outcome-adjusted coverage (6) is the proportion of those in need of a health service who receives the health service according to evidence-based processes of care standards, adheres to provider instructions, and experiences health gains. This approach highlights the importance of each step in the provision of effective health services. In addition, the cascade provides a blueprint to identify bottlenecks in service provision if gaps in effective coverage have been discovered. Yet, individual-level data linking elements of intervention coverage, quality, user-adherence and health outcomes are rarely available in practice. Therefore, most studies have resorted to focusing on certain stages of the cascade; for instance, by estimating service contact coverage nationally from cross-sectional household surveys, and average readiness (for input-adjustment) or average process quality of care (for quality-adjustment) from cross-sectional health facility surveys [99, 100, 101]. However, recent work has highlighted substantial disparities in health-service contact coverage for RMNCH between specific subgroups, such as populations living in different administrative units [65] or socio-economic groups [102]. This suggests that national indicators of effective coverage are poorly-suited to ensure that no subgroup is left behind in the effort of expanding access to high-quality healthcare. Moreover, quality-adjusted coverage has been described as a good proxy for effective coverage if outcome-adjusted coverage measures are not available [103]. Yet, various theoretical frameworks have been used to measure facilities' readiness and process quality [63], leading to non-comparable constructs of quality-adjusted coverage.

In this study, we address these shortcomings by 1) exploring temporal trends in quality-adjusted coverage metrics and disparities between administrative units, and 2) assessing the predictive validity of different quality-adjusted coverage metrics against estimates of child mortality. We develop Bayesian hierarchical models to jointly analyze all publicly available households surveys from the DHS, and health facility surveys from the Service Provision Assessment (SPA) and the Service Delivery Indicators (SDI) surveys, in Senegal, Kenya, and Tanzania - three countries with several rounds of household and facility surveys. We estimate sub-national metrics of quality-adjusted coverage for sick-child care, over time,

using four different metrics corresponding to more or less concise approaches to estimate readiness and process quality of sick-child care. We then consider sub-national estimates of under-5 mortality rates (U5MR) across 118 administrative units - the 45 departments of Senegal, the 47 counties of Kenya, and the 26 regions in Tanzania[1], henceforth referred to as regions. We extract time series of regional-level U5MR between 2010 and 2020 in Senegal, 2000-2020 in Kenya, and 2005-2020 in Tanzania, from a previous study [9]. We explore the association between each of the four metrics of quality-adjusted coverage and under-5 mortality rates at the regional level, using an ecological regression model [81], which includes several regional-level covariates to adjust for differences in environmental risks and preventative health measures coverage, between regions.

Beyond the important research question of determining which construct of quality-adjusted coverage for sick-child care is the most predictive of child health gains, and therefore a good proxy for effective coverage, our findings have practical relevance for future data collection efforts. Health facility assessments are indeed lengthy and costly exercises, and it is essential to critically assess the added-value of more detailed constructs of readiness and quality: are they better at predicting health gains than more succinct constructs?

Methods

Data

Health facility data comes from two large standardized health facility assessment tools, the Service Provision Assessment (SPA) and the Service Delivery Indicators (SDI), while data about child health and health-seeking behaviors were obtained from the DHS, the largest standardized household survey. The SPA and SDI surveys used in this analysis are nationally representative of the formal health sector and based on a stratified survey design - by facility type, and first administrative division, for the SPA, and by urban/rural areas

---

[1]The number of regions in Tanzania increased from 26 to 31, over the study period. Because older SPA and SDI surveys did not contain the GPS coordinates of the facilities, we had to rely on the region indicated in the survey, which correspond to the 2002 subdivisions of Tanzania, which included 26 regions.

and first administrative division in the SDI. DHS surveys are nationally representative and based on a stratified two-stage cluster design – first, enumeration areas are selected from an exhaustive list derived from censuses, then households are sampled within each selected enumeration area. In this analysis, we utilized all publicly available SPA, SDI and DHS surveys in Senegal, Kenya and Tanzania (see Table 4.1). We selected these three countries as several years of both facility and household survey data made it possible to estimate levels and trends in quality-adjusted coverage metrics over time and across sub-national units.

Table 4.1: List of surveys available in Senegal, Kenya and Tanzania

Senegal		Kenya		Tanzania	
Household	Facility	Household	Facility	Household	Facility
DHS 2010-11	SDI 2010[a]	DHS 1998	SPA 1999	DHS 2004-05	SPA 2006
DHS 2012-14[b]	SPA 2012-14[b]	DHS 2003	SPA 2004	DHS 2010	SDI 2010[a]
DHS 2015-16[b]	SPA 2015-16[b]	DHS 2008-09	SPA 2010	DHS 2015-16	SPA 2014
DHS 2017	SPA 2017[b]	DHS 2014	SDI 2012		SDI 2014
DHS 2018	SPA 2018[b]		SDI 2018		SDI 2016
DHS 2019	SPA 2019				

[a]The 2010 SDI surveys in Senegal and Tanzania were pilot studies, which only sampled a small number of public facilities in selected regions of the country, and were therefore excluded from the main analysis.

[b]To ensure that the estimates were geographically representative, DHS data from 2012-13 and 2014-15 were aggregated.

Coverage metrics

We adapted the effective coverage cascade framework presented in Amouzou et al. [53], to focus on quality-adjusted coverage for interventions relating to the management of the

three main causes of death for children under age five: pneumonia, diarrhea, and malaria (see figure 4.1 and table 4.2).

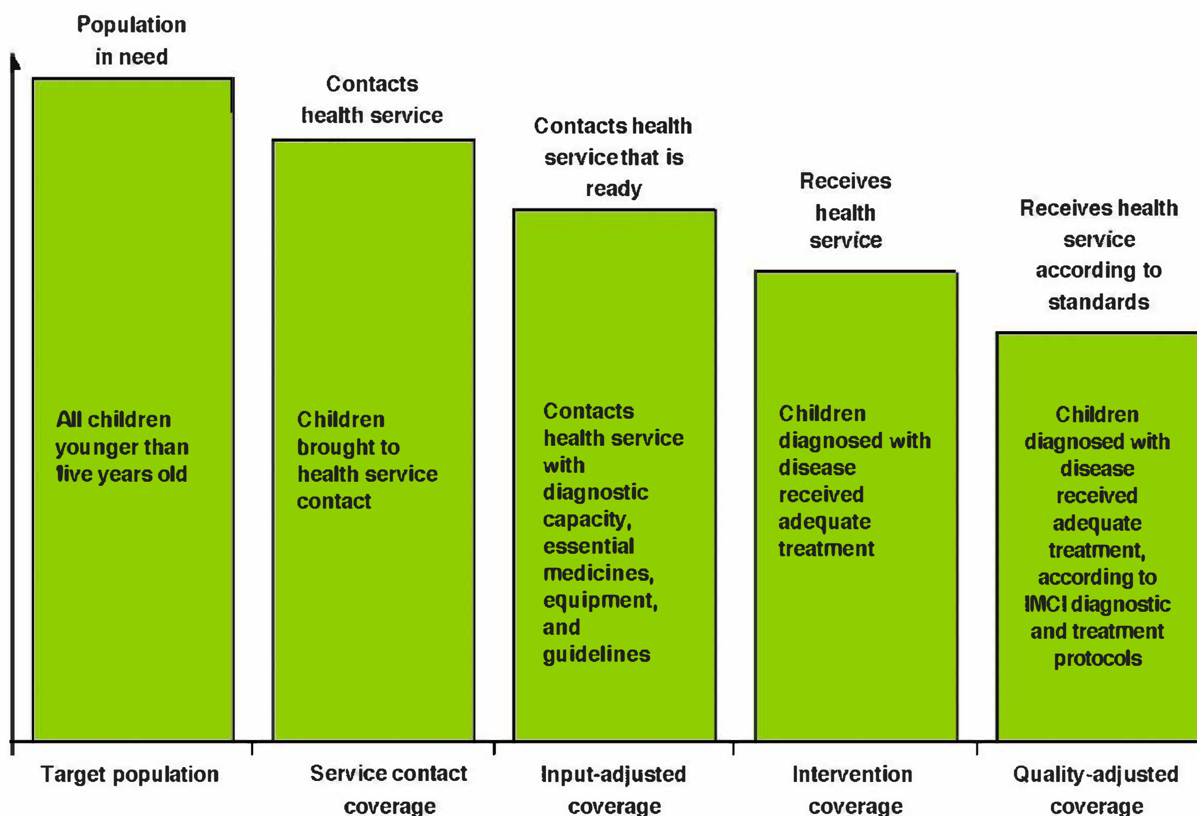


Figure 4.1: Health-service cascade coverage for sick-child care

Service contact coverage, input-adjusted coverage, intervention coverage, and quality-adjusted coverage were calculated sequentially, as described in table 4.2.

We considered four constructs to measure quality-adjusted coverage. In the first approach, we ignore quality of care and focus solely on the proportion of children with symptoms of one of the three conditions that received an appropriate treatment; we therefore calculate quality-adjusted coverage as the product of service contact coverage and the receipt of intervention coverage. The second approach includes the four stages of the cascade, but service readiness and process quality of care are derived from a minimal set of tracer

Table 4.2: Components of quality-adjusted health-service coverage for sick-child care

	Measures			
Cascade	Coverage	× Readiness	× Receipt of complete intervention	× Process quality
Service contact	Children taken to health facility among children under age 5 with diarrhea or ARI or malaria symptoms in the last 2 weeks			
Input-adjusted	Children taken to health facility among children under age 5 with diarrhea or ARI or malaria symptoms in the last 2 weeks	Average readiness[a] across facilities providing sick-child care		
Intervention	Children taken to health facility among children under age 5 with diarrhea or ARI or malaria symptoms in the last 2 weeks	Average readiness across facilities providing sick-child care	Children under age 5 who received appropriate treatment among children diagnosed with diarrhea, pneumonia, malaria at a facility	
Quality-adjusted	Children taken to health facility among children under age 5 with diarrhea or ARI or malaria symptoms in the last 2 weeks	Average readiness across facilities providing sick-child care	Children under age 5 who received appropriate treatment among children diagnosed with diarrhea, pneumonia, malaria at a facility	Average process quality[a] across sick child care facilities

[a]Average readiness and process quality calculated using approaches described in table 3.2

items [61] and diagnostic protocols [62]. The third approach is the same as the second approach, except that service readiness is derived from an extensive list of tracer items. Finally, the fourth approach builds on the third approach, but process quality of care is estimated from a longer list of diagnostic protocols (see table 4.3 for the definition of readiness and process quality of care used in each approach).

Statistical Analyses

Estimating effective coverage metrics and their associated uncertainty

Service contact coverage, readiness, intervention coverage, and process quality of care were modelled separately using a small area estimation approach. In brief, we adapted an existing Bayesian framework, which models direct survey estimates as a function of covariates and space-time components [76]. This model has successfully been used to estimate readiness and process quality of care indicators over time and space from multiple survey tools[2]. Direct estimates and design-based variance estimates were computed using the *survey* package in **R**. We fit the Bayesian hierarchical models using the Integrated Nested Laplace Approximation [32] and the *R-INLA* package [33].

Each quality-adjusted metric was calculated as the product of two (approach 1) or four (approaches 2-4) estimates. To ensure that the quality-adjusted metric reflects the uncertainty associated with each of its components, we drew $k = 1, \ldots, 1000$ posterior samples for all components in region i and year t, and calculated:

$$\theta_{i,t}^{(k)} = \text{service contact}_{i,t}^{(k)} * \text{readiness}_{i,t}^{(k)} * \text{intervention}_{i,t}^{(k)} * \text{process of care}_{i,t}^{(k)}$$

The quality-adjusted metric in region i and year t, and its uncertainty intervals were then derived as the mean, and the 2.5% and 97.5% quantiles.

[2]See previous chapter

Table 4.3: List of tracer items and diagnostic protocols used to derive readiness and process quality of care, by approach

	List of items and protocols included	
Approach	Readiness of child health services	Process quality of care
1	–	–
2	Diagnostic capacity: malaria Essential medicines: ORS for diarrhea and amoxicillin for pneumonia	Provider asked for all three main symptoms: cough or difficulty breathing, diarrhea, fever
3	Diagnostic capacity: malaria Essential medicines: zinc and ORS for diarrhea, amoxicillin for pneumonia, co-trimoxazole, paracetamol Equipment: child and infant scale thermometer, stethoscope Guidelines: IMCI guidelines and growth monitoring guidelines	Provider asked for all three main symptoms: cough or difficulty breathing, diarrhea, fever
4	Essential medicines: zinc and ORS for diarrhea, amoxicillin for pneumonia, co-trimoxazole, paracetamol, antimalarial drugs, Equipment: child and infant scale, thermometer, stethoscope Guidelines: IMCI guidelines and growth monitoring guidelines	IMCI diagnostic protocols: Provider asked for three main symptoms, assessed three danger signs and asked about ear problem took child's temperature, counted respiration, checked skin turgor for dehydration, checked for pallor, looked in child's ear, undressed child to examine, pressed both feet to check for oedema, weighed the child

Ecological regression model

We assessed the association between each of the four quality-adjusted coverage metrics and U5MR over time and across administrative divisions. An ecological regression model enables to pursue both an inferential aim, i.e. obtaining an estimate of the effect of each coverage metric on U5MR, and a predictive aim, i.e. determining which coverage metric best predict out-of-samples mortality rates. We modelled the logit-transform of regional-level U5MR as a linear function of quality-adjusted coverage for sick-child care, covariates and spatially and temporally structured random effects. Covariates utilized in the models include regional-level estimates of diarrhea prevalence [66], and lower-respiratory infection (LRI) incidence [67], to account for differences in risks between regions. Additionally, our model includes regional-level estimates of full immunization coverage (1-dose BCG, 3-dose DTP-HepB-Hib, 3-dose polio, 1-dose measles, and 3-dose pneumococcal vaccines) [104, 105], and access to drinking water and sanitation facilities (WASH) [106], to adjust for potential differences in the coverage of preventative interventions (see the list of covariates, their source and resolution in supplementary table C.1). To avoid potential issues of circularity, we excluded all covariates that were used to estimate U5MR in the initial study[3], such as malaria incidence, travel time to nearest city, or urbanicity [9]. We then obtained a subset of all included covariates by checking for multi-collinearity using the variance inflation factor (VIF) with a threshold of 3. Our ecological regression, which includes spatio-temporal random effects to account for spatial and temporal dependence exhibited by regional-level U5MR, is expressed as:

$$\text{logit}\,(\text{U5MR}_{i,t}) = \beta_0 + \beta_1\theta_{i,t} + X_{i,t}\boldsymbol{\beta} + \gamma_t + \alpha_t + e_i + S_i + \delta_{i,t} \tag{4.1}$$

where $\theta_{i,t}$ is the mean estimate of quality-adjusted coverage in region i and year t, $X_{i,t}$ are the regional-level covariates observed in region i in year t, α_t are temporal random effects modeled

[3] U5MR estimates and their corresponding uncertainty intervals can be downloaded from http://ghdx.healthdata.org/record/ihme-data/lmic-under5-mortality-rate-geospatial-estimates-2000-2017

as a first order random walk, S_i are spatially structured random effects, $\delta_{i,t}$ is a space-time interaction term, and γ_t and e_i are respectively temporal and spatial zero-mean independent random effects. We compared the four quality-adjusted coverage metrics' estimates and their uncertainty intervals obtained from fitting model (4.1).

Predictive validity

We used out-of-sample testing to assess the predictive accuracy of model (4.1). We removed all the observations for region i in year t, and compared the predicted U5MR with the estimate of U5MR reported in the initial study. We compared the predictive accuracy of model (4.1) fitted using the four different quality-adjusted coverage metrics, by comparing coverage, i.e. the frequency at which the model estimates' 95% intervals contained the initial U5MR estimate.

Quality-adjusted coverage metrics comparison

Quality-adjusted coverage metrics were compared based on the two hypotheses that guided our study: 1) evidence of a negative association between quality-adjusted coverage and U5MR, as estimated by our ecological regression model, and 2) evidence of predictive validity of a model including quality-adjusted coverage on U5MR.

Results

Levels and trends in sub-national disparities in quality-adjusted coverage metrics

In Senegal, Kenya, and Tanzania, quality-adjusted coverage for sick child care was low across the four metrics. Figure 4.2 presents time series of predicted mean quality-adjusted coverage (line) and associated uncertainty (shaded area) at the national level, for the four different metrics. Overall, we see that metric 2, 3 and 4 lead to significantly lower mean estimates of quality-adjusted coverage than metric 1, suggesting large gaps in quality and readiness in the three countries. Additionally, we see minimal differences between metrics 2

and 3, indicating that the basic measure of facilities' readiness based on three tracer items lead to almost identical estimates than those obtained using a 12-item measure of readiness. However, considering fifteen rather than 3 diagnostic protocols led to a noticeable drop in the quality-coverage metric 4, compared to metrics 2 and 3. Differences between countries over the study period were not statistically significant, as illustrated by the overlapping uncertainty intervals. National temporal trends are also characterized by substantial uncertainty, but suggest a slight increase in quality-adjusted coverage in Kenya, a stagnation in Senegal, and a decrease in Tanzania, over the study period. The substantial uncertainty both reflects the sparsity of data, and the significant disparities in quality-adjusted coverage between sub-national units. Multiplying sub-national level estimates of service contact coverage, receipt of intervention coverage, basic readiness of child services, and prevalence of the assessment of the three main symptoms leads to our third metric of quality-adjusted coverage mapped for 2020 on figure 4.3. We see uniformly low levels of predicted quality-adjusted coverage, ranging from 8.6% (2.9-17.0) to 27.6% (16.3-41.8) in Senegal, 10% (.01-30.9) to 22% (34.6-44.2) in Kenya, and 2% (.01-11.9) to 14% (1.1-40.1) in Tanzania. Quality-adjusted coverage seems marginally higher than the national average in the southeastern regions of Senegal, in the northwestern regions of Kenya, and in the north of Tanzania.

Drivers of disparities and specific gaps

We present in figure 4.4 a decomposition of quality-adjusted coverage along the two dimensions of service contact coverage against intervention coverage, for each country, in 2010 and 2020, in order to identify health system priorities along these two dimensions. In Senegal, receipt of intervention coverage is high and shows little variability, with most sub-national units in the upper quadrant of the figure, but service contact coverage is low and varies significantly. In Kenya, most sub-national units in 2020 achieve high service and intervention coverage in the upper right quadrant, i.e above 50%, which suggests important improvements since 2010, when receipt of intervention coverage was below 50% for most sub-national units. In Tanzania, all sub-national units are comprised on the upper and lower right quadrant

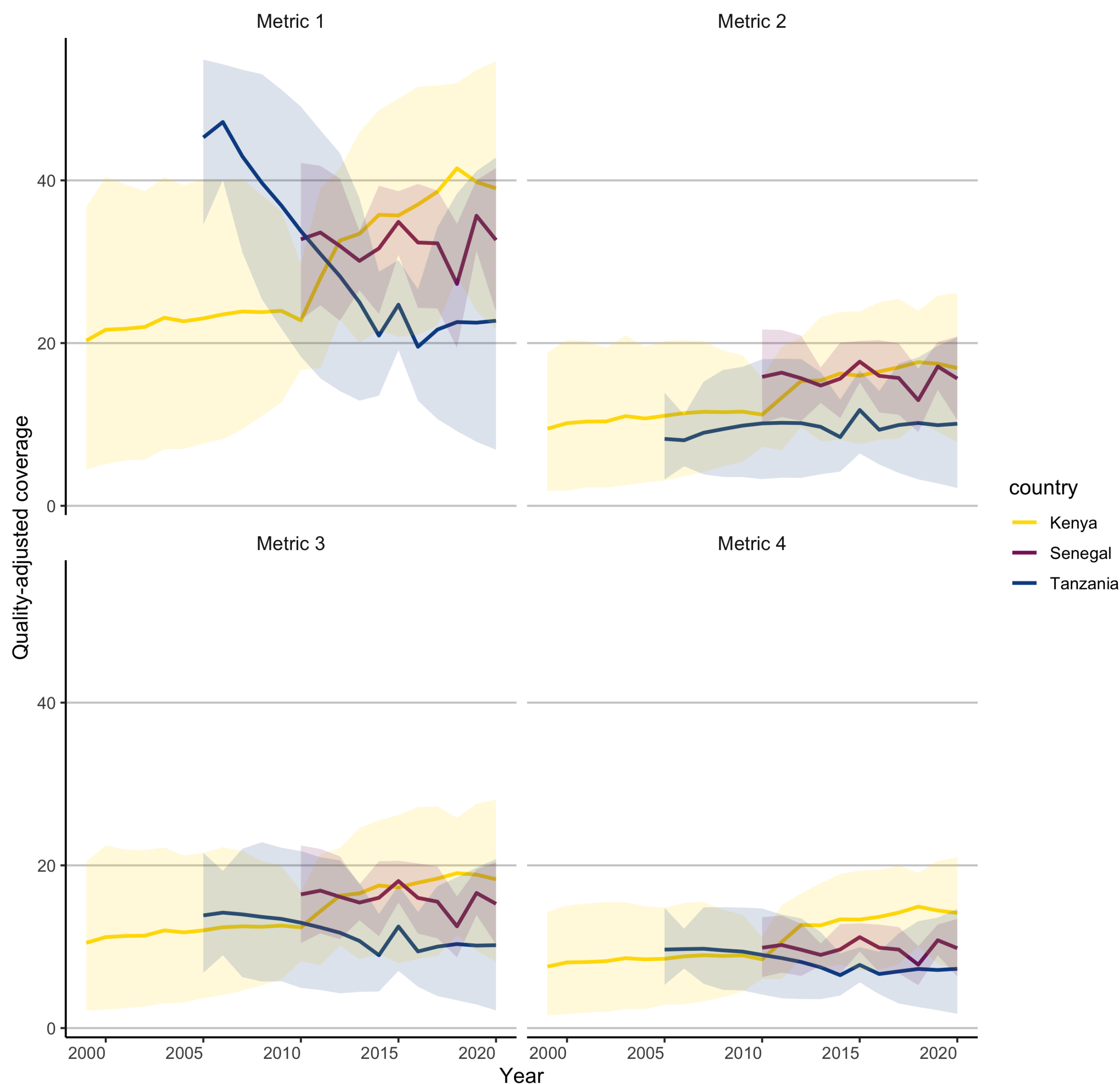


Figure 4.2: Metrics of quality-adjusted coverage over time in Senegal, Kenya and Tanzania

of figure, suggesting levels of service contact coverage above 50%; receipt of intervention coverage is however less than 50% in most sub-national units, over the study period, and for all sub-national units in 2020. Supplementary figure C.1 adds a spatial lens by mapping

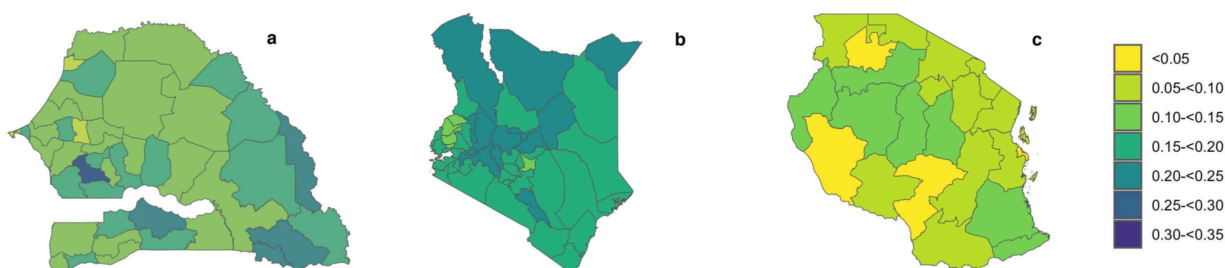


Figure 4.3: Mean estimates of quality-adjusted coverage metric 3 in 2020, in Senegal (a), Kenya (b) and Tanzania (c)

the joint distribution of service contact and receipt of intervention coverage; administrative units are grouped by quantiles to compare their relative performances on these two cascade stages and identify priority areas. Figure C.1**a-b** suggests that Senegalese sub-national units with lowest service contact coverage are located in the southern and southeastern parts of the country. In Kenya, highest coverage of both service contact and receipt of intervention in the center of the country, lower service contact coverage in the Northeastern part of the country, and lower intervention coverage in the lake region and the southeastern part of the country (see figure C.1**c-d**). In Tanzania, figure C.1**e-f** suggests consistent spatial patterns between 2010 and 2020, with higher receipt of intervention coverage but lower service contact coverage in the northwestern regions of the country, higher service contact coverage but lower intervention coverage in the south, and higher levels of both service and intervention coverage in the eastern regions of the country.

Association between quality-adjusted coverage metrics and child mortality

We have fitted the model specified in equation (4.1), with the four different metrics described in table 4.3 for quality-adjusted coverage. Point estimates of the fixed effects, along with their associated standard deviation (SD), for the four models, are reported in table 4.4. In general, our results match our expectations: prevalence of diarrhea and incidence of LRI

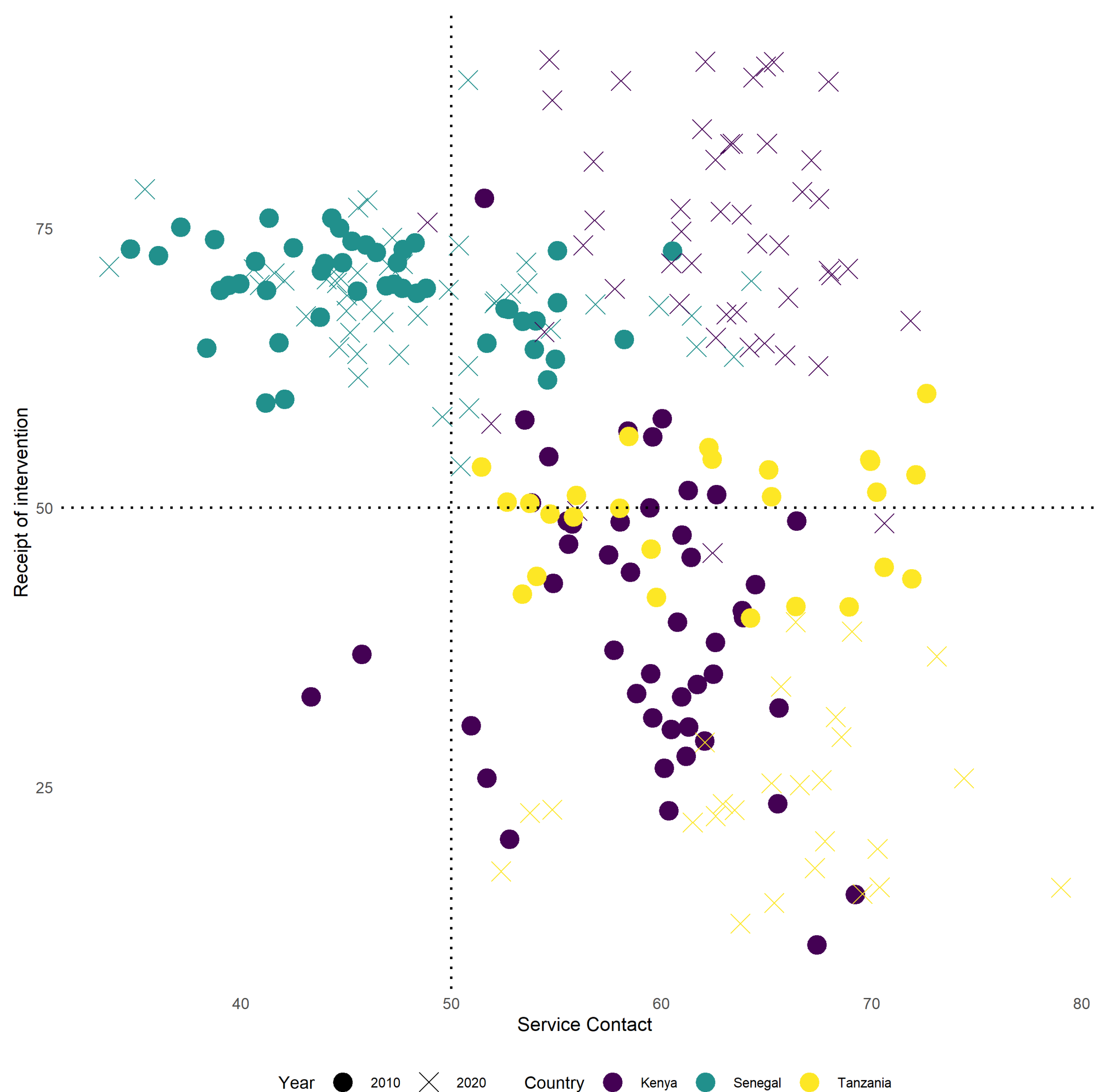


Figure 4.4: Service contact and receipt of intervention coverage in Senegal, Kenya and Tanzania, by sub-national units in 2010 and 2020

are positively associated with under-five mortality, while coverage of the different vaccines are negatively associated with under-five mortality. Our main predictor of interest, quality-

adjusted coverage, was found to be negatively associated with under-five mortality, in all four models, suggesting a protective effect at the ecological level. It is worth noting that as the quality-adjusted coverage incorporates more elements of quality, its estimated mean effect becomes more strongly negative, but the width of its credible interval increases as well: $\beta_1 = -.82$ (95% CI 1.66 - 0.02) in model 1, but $\beta_1 = -1.945$ (-4.105 - 0.21) in model 4. Fixed effects associated with the country variable were small and non-significant. Interaction terms between the country and the quality-adjusted coverage variables were almost statistically significant in Senegal and Tanzania. This suggests that, for a given level of quality-adjusted coverage, mortality rates were higher in Senegal and Tanzania than in Kenya, on average. Visualizations of the temporal and spatial random effects are provided in supplementary section C.3. U5M has decreased significantly over our study period, especially in the first five years (see figure C.2). Spatially structure random effects indicated positive spatial residuals in the southeastern parts of Senegal, the lake region of Kenya and the southern regions of Tanzania (see figure C.3) - three areas where malaria, whose incidence was not included in the model to avoid circularity, is endemic.

Metrics of quality-adjusted coverage comparison

Higher quality-adjusted coverage means a higher proportion of children with recent illnesses who received appropriate treatment at a health facility, and should therefore not be positively associated with U5MR. We see with table 4.4 that none of the four quality-adjusted coverage metrics were disqualified based on this criteria, as all were negatively associated with U5MR. We further tested the sign of this association by visualizing the posterior densities of each metric estimated effect. Supplementary figure C.4 shows the posterior distributions of the association between quality-adjusted coverage and U5MR; the “Bayesian p-value", or probability that the association is negative, corresponds to the area of the density that below 0, which ranges from 93.9% with metric 1 to 97.7% with metric 3.

Quality-adjusted coverage metrics should help predict under-five mortality rates; we used two criteria to assess the four metrics' ability to estimate U5M - the mean-squared errors

Table 4.4: Posterior means of the fixed effect coefficients for the four models

Predictors	Metric 1		Metric 2		Metric 3		Metric 4	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
(Intercept)	-2.564	(.134)	-2.629	(.126)	-2.555	(.140)	-2.611	(.119)
Quality-adjusted coverage	-0.824	(.420)	-1.418	(.935)	-1.787	(.904)	-1.945	(1.086)
Country								
Kenya	*Ref.*	*Ref.*	*Ref.*	*Ref.*				
Senegal	-0.577	(0.337)	-0.543	(0.251)	-0.615	(0.261)	-0.506	(.241)
Tanzania	0.033	(.227)	0.129	(.223)	0.053	(.223)	0.088	(.219)
LRI Incidence	0.048	(.040)	0.062	(.040)	0.054	(.040)	0.057	(.040)
Diarrhea Prevalence	0.198	(.045)	0.192	(.045)	0.192	(.045)	0.197	(.044)
MCV1 coverage[a]	-0.040	(.034)	-0.037	(.034)	-0.035	(.034)	-0.036	(.034)
PCV3 coverage[b]	-0.026	(.050)	-0.020	(.050)	-0.022	(.050)	-0.022	(.050)
Polio3 coverage[b]	-0.020	(.033)	-0.010	(.033)	-0.014	(.033)	-0.017	(.033)
Hib3 coverage[b]	-0.026	(.039)	-0.046	(.038)	-0.041	(.038)	-0.037	(.037)
Interaction country* quality-adjusted coverage								
Kenya*coverage	*Ref.*	*Ref.*	*Ref.*	*Ref.*				
Senegal*coverage	1.462	(1.021)	3.023	(1.543)	3.291	(1.536)	3.700	(2.116)
Tanzania*coverage	1.693	(.710)	3.445	(1.840)	3.636	(1.596)	4.879	(2.283)

[a] % of children under one year of age who have received at least one dose of measles-containing vaccine (MCV1) in a given year.

[b] % of children ages 12–23 months who received three doses three doses of pneumococcal conjugate vaccine (PCV3), polio vaccine (Polio3), and Haemophilus influenzae type B vaccine (Hib3), in a given year.

and the leave-one-out predictions' 95% coverage. We present the results in table 4.5: metrics 1 and 3 provided predictions that were the closest to the initial estimates, as indicated by the mean-squared errors.

Table 4.5: Performances of the four models using the four different quality-adjusted metrics, assessed using mean squared errors (MSE) and spatial leave-one-out predictions of U5MR in the 118 sub-national units across Senegal, Kenya and Tanzania

Criteria	Metric 1	Metric 2	Metric 3	Metric 4
MSE	3.1×10^{-7}	6.4×10^{-7}	3.2×10^{-7}	7.3×10^{-7}
Coverage (%)	98	88	95	92

Discussion

Improvements in the management of sick children will be key to sustain the reduction in the burden of child mortality recorded in LMICs, in the last few decades. Quality-adjusted coverage for sick-child care, a measure of the proportion of children with symptoms of pneumonia, diarrhea, or malaria, that were diagnosed and treated at a health facility according to evidence-based guidelines, is a critical metric to identify and resolve gaps in the availability, readiness and quality of child health services. This study is, to our knowledge, the first model-based approach to estimating quality-adjusted coverage at the sub-national level and over time, using household and health facility surveys. Overall, we estimated low levels of quality-adjusted coverage in the three countries over the study period, and predicted for 2020 a coverage of 15.3%, 18.3%, and 10.2% in Senegal, Kenya, and Tanzania, respectively - which supports and broadens previous findings. Riese and colleagues reported national cross-sectional estimates of quality-adjusted coverage of 8.7% in Senegal in 2018, and 11.5% in Tanzania in 2015, [107]. Using similar data sources but a narrower metric,

Leslie et al. reported a coverage of 22.9% in Kenya in 2010, 11.7% in Senegal in 2014 and 15.6% Tanzania in 2015 [100]. Our spatial analyses added evidence of substantial and persistent gaps at every stage of the cascade- in service contact coverage, readiness of services, receipt of intervention coverage and process quality of care. In particular, we estimated low levels of service intervention coverage across sub-national units in Senegal, and low levels of receipt of intervention coverage across units of Tanzania. In Kenya, we predicted moderate but improving levels of both service and receipt of intervention. Our spatio-temporal estimates are directly relevant for assessing the impact of past interventions. For instance, in Senegal, where USAID has defined seven "regions of concentration", including the southeastern regions of Kedougou and Tambacounda, with targeted direct funding between 2016 and 2021 dedicated to improving access and quality to RMNCH services. Our results suggest that several of these regions have recorded large gains in quality-adjusted coverage for sick-child care between 2016 and 2020.

Combining our spatio-temporal estimates with time series of sub-national indicators of vaccination coverage, and environmental risks, we investigated the association between quality-adjusted coverage, measured using four different approaches, and under-five mortality rates. We found evidence of a protective effect of quality-adjusted coverage, regardless of the metric used, against child mortality, at the ecological level, although not statistically significant. We found evidence of a positive interaction between quality-adjusted coverage and country effects in Senegal and Tanzania. This interaction may reflect some country-specific policies not accounted for in the model; for instance, Senegalese southeastern regions, which have experienced the highest under-five mortality rates over the study period, are recipients of targeted international interventions [83], and have thus experienced large gains in service contact coverage. This may also reflect the different mix of diseases that contribute to child mortality in the three countries, or the differential benefits of health interventions not accounted for in our model. For instance, pediatric HIV, which accounts for a much larger share of deaths in Kenya than Tanzania or Senegal, has decreased over our study period with the progress of mother-to-child transmission efforts [108]. Similarly, the uptake

of insecticide-treated bed-nets has been significantly higher in Kenya than in Senegal or to a lesser extent Tanzania [109].

We constructed four different quality-adjusted coverage metrics from recent frameworks, developed to measure healthcare quality and effective coverage. Our approach to comparing the four metrics was primarily data-driven: our ecological regression provided a framework to assess their performances in predicting U5MR, at the ecological level. Similarly to Riese and colleagues [107], we found that metrics using more concise or more expanded constructs led to highly-correlated estimates of quality-adjusted coverage. Our results further indicated that metrics using a succinct list of tracer items and diagnostic protocols to measure readiness and process quality of care, performed, at least equally as more complex metrics in predicting under-five mortality rates. These results seem to confirm previous critical assessments of frameworks such as the World Health Organization's SARA index, which are used to design facility surveys, and lead to the collection of extensive, but partially redundant, information on health services readiness and quality, while ignoring other important aspects of healthcare quality [58]. More generally, we tried to highlight the fundamental tension in healthcare quality measurement between producing quantitative metrics meant to be evidence for policy makers, and simultaneously critically assessing the limitations of these metrics. This study epitomizes this tension: we created four metrics to measure a complex concept, quality-adjusted coverage, and used them to evidence sub-national disparities, before proof-testing these constructs, with a regression framework, to contextualize and connect them to other important health indicators.

Our work is not without several important limitations. First, to construct quality-adjusted coverage metrics, we linked household and facility-level indicators at the sub-national level- departments, counties and regions, which precludes us from identifying disparities at a smaller scale, between cities or neighborhoods. Additionally, this administrative linkage introduces measurement errors as it ignores the possibility that households seek care in an adjacent administrative unit [110]. However, alternative techniques, such as direct linkage based on distance or travel time, would have introduced significant mis-classification

errors [93]. Second, our measures of quality-adjusted coverage for sick-child care is restricted to sick children who sought care at a formal, public or private, health facility; the SPA and SDI, from which we derived intervention coverage, readiness of services and process quality of care, do not survey non-formal healthcare providers. However, we know that non-formal providers, such as traditional healers, or private pharmacies can be involved in the management of sick children; for instance, in Senegal, up to 40% of parents will first seek diarrhea treatment for their children at private pharmacies [111]. Therefore, our estimates could be underestimating the true level of coverage. Third, to calculate input-adjusted and quality-adjusted coverage metrics, we estimated facilities' readiness and process quality of care, using WHO guidelines- the SARA and IMCI, to identify relevant items and diagnostic protocols. Yet, standardized guidelines may only capture some aspects of service provision, which is inherently multi-faceted and therefore challenging to assess [86]. Fourth, our regression analysis of the association between quality-adjusted coverage and U5MR was done at the ecological study, which leads to a substantial information loss: key individual-level determinants of health-seeking behaviours and child survival, such as mother's education or socio-economic status, and child's gender, age or order, could not be included in the analyses [10]. Additional potential omitted variables include important predictors of child mortality excluded to avoid circularity, such as malaria incidence, or because fine-scale estimates were not available - for instance, pediatric HIV. Further threats to the consistency of our estimates relate to the assumption that the effect of curative services (quality-adjusted coverage) and preventive services (immunization coverage) on U5MR was additive. Given the documented strong association between delivering these two interventions, alternative functional forms (such as multiplicative or interaction terms) could have been more appropriate, but were ruled out for simplicity of interpretation. Our results should therefore be interpreted cautiously as evidence of association, rather than of causal inference. Lastly, the outcome of our ecological regression, as well as most predictors, including the quality-adjusted coverage metrics, are statistical estimates with associated uncertainty. Yet, we used their mean estimates as error-free, which can introduce biases in parameters and confidence intervals

estimates, as well as reduce the power for detecting associations between variables [112]. While ignoring measurement error in explanatory variables is the most common approach in the global health literature, future work could incorporate recent development in statistics, to account for errors in ecological covariates measurement [113], and therefore improve estimates of impact. Other possible extensions of our work toward inferential approaches include an individual-level regression analyses of the association between quality-adjusted coverage and child survival. Two recent studies have investigated the relationship between healthcare quality and child survival, but lacked information on service and intervention coverage [70, 94]. Furthermore, additional studies are needed to better understand the impact of health systems' organizational features on child mortality; in particular, beyond individual providers' adequate diagnostic and treatment of children sickness, appropriate and timely referral practices or the lack thereof often have dramatic consequences on sick children survival [114]. Such work would require integrating new data sources including routine health information systems, which would be more suited than cross-sectional facility surveys to track children patients' path through different levels of care, and identify bottlenecks.

Conclusion

Quality-adjusted coverage for sick-child care is a critical metric to identify and resolve gaps in the availability, readiness and quality of child health services. Our results reveal large and persistent within-country inequalities in quality-adjusted coverage, which may undermine the sustained progress in reducing the burden of child mortality. The mapping of estimates of service contact coverage, receipt of intervention coverage, readiness and process quality of care, at fine spatial resolution, can be used to identify and resolve local gaps in quality-adjusted coverage and ensure that the progress toward universal health coverage leave no one behind.

Chapter 5

CONCLUSIONS

Summary Aims 1-3

Tuberculosis can be treated, prevented, and cured with effective life-saving therapies [40]. Yet, up to 35% of people living with tuberculosis are not diagnosed, notified to national tuberculosis control programs, and treated. Achieving the UN's and WHO's ambitious strategy to "End TB" will require to improve access to high-quality tuberculosis services, especially in high-burden countries [115], which includes ensuring that diagnostic and treatment capacity is located where the need is the greatest. In Aim 1, we addressed the need to understand sub-national variations in tuberculosis prevalence, and prevalence-to-notification ratio, which has important implications for infection control, including determining local need for tuberculosis treatments, and identifying gaps in routine surveillance activities, in Bangladesh, a high-burden country. Specifically, we adapted a Bayesian spatial framework to jointly analyze national prevalence and case notification data from Bangladesh National Tuberculosis Control Program, and generated estimates for second-level administrative subdivisions (districts). The results of this work indicated significant sub-national variations in tuberculosis prevalence in Bangladesh, with high-burden districts in the northern part of the country. Importantly, this work provided further evidence of substantial gaps in the diagnosis and reporting of tuberculosis cases in two clusters of districts, identified in previous studies [25]. Our study presented methods and estimates that can help guide the efforts required to close the gaps in diagnostic, prevention, and treatment in Bangladesh, and other high-burden countries with similar data sources.

Poor-quality care is a major barrier to health gains in most low- to middle-income countries. Attempts to measure health services quality, to date, have focused on cross-sectional,

national studies, which overlooks temporal trends and sub-national inequities in healthcare quality. Characterizing temporal trends and sub-national disparities in readiness and process quality metrics, in three low- to middle-income countries - in Senegal, Kenya and Tanzania, using health facility surveys, was the focus of the second Aim of this dissertation. We had a particular interest in exploring the possible gains of a statistical framework supplementing direct survey estimates with model-based predictions, building on correlation in time and space and auxiliary information. In particular, we thought that such an approach could provide time series of readiness and process quality estimates, at a finer spatial resolution than what the surveys were powered to do. We found an overall good out-of-sample predictive validity of our models, with relatively high coverage and low mean-squared errors, across metrics and countries. Our sub-national estimates evidenced overall gaps in readiness and process quality metrics in Senegal, Kenya and Tanzania, with substantial within-country disparities. Our study introduced new statistical methods that enhance the measurements of actionable quality metrics, at a sub-national resolution that can help inform programmatic decision-making and enhance equity in availability of high-quality services.

Finally, Aim 3 sought to address an important challenge in healthcare quality measurement research; identifying relevant metrics to monitor access to health services of sufficient quality. Quality-adjusted coverage is defined as the proportion of the population in need of a health service that receives appropriate treatment according to standards, and as such, has been described as an adequate metric to track countries' progress toward universal health coverage. We described temporal trends and sub-national disparities in quality-adjusted coverage for sick-child care in three low- to middle-income countries; Senegal, Kenya and Tanzania. We estimated four different metrics of quality-adjusted coverage, corresponding to more or less concise approaches to estimate readiness and process quality of sick-child care, from households and health facility data, and compared their performance in predicting under-five mortality rates, over time, and at sub-national levels, using an ecological regression framework. We found uniformly low-levels of quality-adjusted coverage across the sub-national units of the three countries, with substantial spatial disparities along the

cascade of service provision (service coverage, intervention coverage). We reported a negative association between quality-adjusted coverage and under-five mortality rates at the ecological-level. The results of this aim were partially inconclusive, as we found no evidence of one quality-adjusted coverage metric out-performing the others. However, the result that a succinct list of tracer items and diagnostic protocols to measure readiness and process quality of care, performed, at least equally as more complex metrics in predicting under-five mortality rates may have implications for data collection efforts, and confirm previous studies suggesting that current tools and frameworks used to estimate readiness and process quality may measure redundant or irrelevant aspects of service provision while ignoring other important aspects [58].

Future works

This dissertation represents an additional contribution to a growing body of work on service provision in low- to middle-income countries. Following Tanahashi's seminal typology [98], health service coverage can be broken into a cascade of key "stages". Measures of service contact coverage have become available in LMICs with the increased availability of household surveys, such as the Demographic and Health Survey. Service contact estimates are calculated as the proportion of household members in need of a health service that sought care with a health professional. The uptake of health facility surveys, such as the Service Provision Assessment and Service Delivery Indicators, has provided measures of the availability coverage, and insights into the readiness and process quality of care dispensed in formal health structures. Finally, estimates of accessibility coverage have blossomed in recent years with the enhanced use of GIS techniques in public health [48, 49].
Our dissertation proposed to build on this stream of work, by supplementing direct survey estimates of the cascade with small area estimation methods, to describe temporal trends and provide estimates at sub-national resolutions. By doing so, we were able to investigate the question of equity in health service coverage, in the countries studied. We were also able to test empirically the consistency of commonly-used constructs of health services readiness

and process quality of care. Nevertheless, important gaps remain. For instance, none of this dissertation focused on acceptability coverage, which corresponds to people's willingness to use accessible services and is influenced by factors, such as affordability, ease of use, dignity, privacy, or non-discrimination. However, these factors are integral to the equitable provision of high-quality services. Our Aim 1 would particularly benefit from the lens of acceptability, in part as tuberculosis is characterized by significant gender differences in accessing care that might be related to both financial barriers and stigma [116].

With regard to Aim 2, future work should focus on trying to critically assess the consistency and composition of composite indicators used to describe facilities' readiness and process quality. First, calculated as arithmetic means of survey items, these composite indicators may over-represent easy to measure areas of care, fail to represent the relative relevance to quality of each item, and fail to reproduce the uncertainty associated with the measurement of each item [85]. Latent variable models, including item factor analyses methods, could provide an approach that would down-weight redundant items and account for correlation between them, to supply a summary metric of readiness or process quality with uncertainty. Second, indicators of readiness and process quality of care are derived from assessment frameworks and clinical charts that can be poorly suited to local realities. For instance, in aim 2, we found in Senegal and Tanzania, evidence of seasonality in the frequency of the assessment of fever in sick children, which reflects seasonality in the risk of malaria, suggesting an adaptive application of the guidelines by providers facing a high caseload. More broadly, going through these guidelines may resemble a box-ticking exercise if the provider has already diagnosed the disease. Future research could study further the reasons for the extensive gaps in adherence to evidence-based guidelines and ways to resolve them.

With regard to aim 3, individual-level regression analyses of the association between quality-adjusted coverage and child survival could expand our evidence of a protective effect. In particular, individual-level analyses would allow to include key individual-level determinants of health-seeking behaviours and child survival, such as mother's education or

socio-economic status, and child's gender, age or order. Assessing the relative importance of different modifiable factors, such as immunization status, utilization of quality health-care, access to improved water sources and sanitation, and wealth, on individual child survival could have tremendous implications in terms of policy-making and prioritization of interventions.

BIBLIOGRAPHY

[1] Rafael Lozano, Nancy Fullman, John Everett Mumford, et al. Measuring universal health coverage based on an index of effective coverage of health services in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, 396(10258):1250–1284, October 2020. Publisher: Elsevier.

[2] Margaret E Kruk, Anna D Gage, Naima T Joseph, et al. Mortality due to low-quality health systems in the universal health coverage era: a systematic analysis of amenable deaths in 137 countries. *The Lancet*, 392(10160):2203–2212, November 2018.

[3] Margaret E Kruk, Anna D Gage, Catherine Arsenault, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution. *The Lancet Global Health*, 6(11):e1196–e1252, November 2018.

[4] Nancy Fullman, Jamal Yearwood, Solomon M Abay, et al. Measuring performance on the Healthcare Access and Quality Index for 195 countries and territories and selected subnational locations: a systematic analysis from the Global Burden of Disease Study 2016. *The Lancet*, 391(10136):2236–2271, June 2018.

[5] Kathryn O'Neill, Marina Takane, Ashley Sheffel, et al. Monitoring service delivery for universal health coverage: the Service Availability and Readiness Assessment. *Bull World Health Organ*, 91(12):923–931, December 2013. Number: 12.

[6] Sandra Alba, Ente Rood, Mirjam I Bakker, et al. Development and validation of a predictive ecological model for TB prevalence. *Int J Epidemiol*, 47(5):1645–1657, October 2018.

[7] Erlyn K Macarayan, Anna D Gage, Svetlana V Doubova, et al. Assessment of quality of primary care with facility surveys: a descriptive analysis in ten low-income and middle-income countries. *The Lancet Global Health*, 6(11):e1176–e1185, November 2018.

[8] Zehang Li, Yuan Hsiao, Jessica Godwin, and others with support from the United Nations Inter-agency Group for Child Mortality Estimation and its technical advisory group. Changes in the spatial distribution of the under-five mortality rate: Small-area analysis of 122 DHS surveys in 262 subregions of 35 countries in Africa. *PLoS ONE*, 14(1):e0210645, January 2019.

[9] Roy Burstein, Nathaniel J. Henry, Michael L. Collison, et al. Mapping 123 million neonatal, infant and child deaths between 2000 and 2017. *Nature*, 574(7778):353–358, October 2019.

[10] Deborah Balk, Thomas Pullum, Adam Storeygard, et al. A spatial analysis of childhood mortality in West Africa. *Population, Space and Place*, 10(3):175–216, 2004. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/psp.328.

[11] World Health Organization. Global Tuberculosis Report 2020. Technical report, World Health Organization, Geneva, 2020.

[12] World Health Organization. End TB Strategy. Technical report, World Health Organization, Geneva, 2015.

[13] United Nations. Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development. Springer Publishing Company, New York, NY, June 2017.

[14] World Health Organization, editor. *Understanding and using tuberculosis data*. World Health Organization, Geneva, Switzerland, 2014. OCLC: ocn897868764.

[15] Hmwe Hmwe Kyu, Emilie R Maddison, Nathaniel J Henry, et al. Global, regional, and national burden of tuberculosis, 1990–2016: results from the Global Burden of Diseases, Injuries, and Risk Factors 2016 Study. *The Lancet Infectious Diseases*, 18(12):1329–1349, December 2018.

[16] P Glaziou, Peter J Dodd, A Dean, et al. Methods used by WHO to estimate the global burden of TB disease. Technical report, World Health Organization, October 2020.

[17] K. Zaman, S. Hossain, S. Banu, et al. Prevalence of smear-positive tuberculosis in persons aged $\geq$ 15 years in Bangladesh: results from a national survey, 2007-2009. *Epidemiol Infect*, 140(6):1018–1027, June 2012.

[18] National Tuberculosis Control program. Tuberculosis Prevalence Survey Report Bangladesh 2015-2016. Technical report, National Tuberculosis Control program, 2017.

[19] Xin-Xu Li, Li-Xia Wang, Hui Zhang, et al. Spatial variations of pulmonary tuberculosis prevalence co-impacted by socio-economic and geographic factors in People's Republic of China, 2010. *BMC Public Health*, 14(1):257, March 2014.

[20] Kiesha Prem, Sok Heng Pheng, Alvin Kuo Jing Teo, et al. Spatial and temporal projections of the prevalence of active tuberculosis in Cambodia. *BMJ Global Health*, 4(1):e001083, January 2019. Publisher: BMJ Specialist Journals Section: Research.

[21] KIT. Hack TB, Centre for Applied Spatial Epidemiology, Pakistan's National TB Control Program, Stop TB partnership, 2019.

[22] Kefyalew Addis Alene, Zeleke Alebachew Wagaw, and Archie C. A. Clements. Mapping tuberculosis prevalence in Ethiopia: protocol for a geospatial meta-analysis. *BMJ Open*, 10(5):e034704, May 2020. Publisher: British Medical Journal Publishing Group Section: Epidemiology.

[23] C. G. Parwati, M. N. Farid, H. S. Nasution, et al. Estimation of subnational tuberculosis burden: generation and application of a new tool in Indonesia. *The International Journal of Tuberculosis and Lung Disease*, 24(2):250–257, February 2020.

[24] Jennifer M. Ross, Nathaniel J. Henry, Laura A. Dwyer-Lindgren, et al. Progress toward eliminating TB and HIV deaths in Brazil, 2001–2015: a spatial assessment. *BMC Med*, 16(1):144, December 2018.

[25] Ente Rood, Ahmadul Khan, Pronab Modak, et al. A Spatial Analysis Framework to Monitor and Accelerate Progress towards SDG 3 to End TB in Bangladesh. *IJGI*, 8(1):14, December 2018.

[26] Margo van Gurp, Ente Rood, Razia Fatima, et al. Finding gaps in TB notifications: spatial analysis of geographical patterns of TB notifications, associations with TB program efforts and social determinants of TB risk in Bangladesh, Nepal and Pakistan. *BMC Infectious Diseases*, 20(1):490, July 2020.

[27] Mesay Hailu Dangisso, Daniel Gemechu Datiko, and Bernt Lindtjørn. Accessibility to tuberculosis control services and tuberculosis programme performance in southern Ethiopia. *Global Health Action*, 8(1):29443, December 2015. Publisher: Taylor & Francis _eprint: https://doi.org/10.3402/gha.v8.29443.

[28] Cici Chen, Jon Wakefield, and Thomas Lumely. The use of sampling weights in Bayesian hierarchical models for small area estimation. *Spatial and Spatio-temporal Epidemiology*, 11:33–43, October 2014.

[29] Laina Mercer, Jon Wakefield, Cici Chen, et al. A comparison of spatial smoothing methods for small area estimation with sampling weights. *Spatial Statistics*, 8:69–85, May 2014.

[30] Theo Vos, Stephen S. Lim, C Abbafati, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, 396(10258):1204–1222, October 2020. Publisher: Elsevier.

[31] Robert E. Fay and Roger A. Herriot. Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, 74(366a):269–277, June 1979. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.1979.10482505.

[32] Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392, 2009. _eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2008.00700.x.

[33] Finn Lindgren and Håvard Rue. Bayesian Spatial Modelling with R-INLA. *Journal of Statistical Software*, 63(1):1–25, February 2015. Number: 1.

[34] Daniel Marguari, Carmelia Basri, Wenita Indrasari, et al. Social Barriers to Accessing Quality TB Service: TB Key Populations, Legal Environment and Gender Assessment. page 95.

[35] C. Dye and B. G. Williams. The Population Dynamics and Control of Tuberculosis. *Science*, 328(5980):856–861, May 2010.

[36] Madeeha Laghari, Syed Azhar Syed Sulaiman, Amer Hayat Khan, et al. Contact screening and risk factors for TB among the household contact of children with active TB: a way to find source case and new TB cases. *BMC Public Health*, 19, September 2019.

[37] Courtney M Yuen, Farhana Amanullah, Ashwin Dharmadhikari, et al. Turning off the tap: stopping tuberculosis transmission through active case-finding and prompt effective treatment. *The Lancet*, 386(10010):2334–2343, December 2015.

[38] S. Hossain, K. Zaman, S. Banu, et al. Tuberculin survey in Bangladesh, 2007–2009: prevalence of tuberculous infection and implications for TB control. *The International Journal of Tuberculosis and Lung Disease*, 17(10):1267–1272, October 2013.

[39] H. Esmail, C. E. Barry, D. B. Young, et al. The ongoing challenge of latent tuberculosis. *Philos Trans R Soc Lond B Biol Sci*, 369(1645), June 2014.

[40] Michael J. A. Reid, Nimalan Arinaminpathy, Amy Bloom, et al. Building a tuberculosis-free world: The Lancet Commission on tuberculosis. *The Lancet*, 393(10178):1331–1384, March 2019. Publisher: Elsevier.

[41] Danielle Cazabon, Hannah Alsdurf, Srinath Satyanarayana, et al. Quality of tuberculosis care in high burden countries: the urgent need to address gaps in the care cascade. *International Journal of Infectious Diseases*, 56:111–116, March 2017.

[42] Shahed Hossain, Mohammad Abdul Quaiyum, Khalequ Zaman, et al. Socio Economic Position in TB Prevalence and Access to Services: Results from a Population Prevalence Survey and a Facility-Based Survey in Bangladesh. *PLoS ONE*, 7(9):e44980, September 2012.

[43] TB service disruption simulator.

[44] Ivor Langley, Hsien-Ho Lin, Saidi Egwaga, et al. Assessment of the patient, health system, and population effects of Xpert MTB/RIF and alternative diagnostics for tuberculosis in Tanzania: an integrated modelling approach. *The Lancet Global Health*, 2(10):e581–e591, October 2014. Number: 10.

[45] James M Trauer, Peter J Dodd, M Gabriela M Gomes, et al. The Importance of Heterogeneity to the Epidemiology of Tuberculosis. *Clinical Infectious Diseases*, 69(1):159–166, June 2019.

[46] Kirsten E. Wiens, Lauren P. Woyczynski, Jorge R. Ledesma, et al. Global variation in bacterial strains that cause tuberculosis disease: a systematic review and meta-analysis. *BMC Medicine*, 16(1):196, October 2018.

[47] Paul O. Ouma, Joseph Maina, Pamela N. Thuranira, et al. Access to emergency hospital care provided by the public sector in sub-Saharan Africa in 2015: a geocoded inventory and spatial analysis. *The Lancet Global Health*, 6(3):e342–e350, March 2018. Publisher: Elsevier.

[48] A. S. Wigley, N. Tejedor-Garavito, V. Alegana, et al. Measuring the availability and geographical accessibility of maternal health services across sub-Saharan Africa. *BMC Medicine*, 18(1):237, September 2020.

[49] D. J. Weiss, A. Nelson, C. A. Vargas-Ruiz, et al. Global maps of travel time to healthcare facilities. *Nature Medicine*, 26(12):1835–1838, December 2020. Number: 12 Publisher: Nature Publishing Group.

[50] Margaret E Kruk, Hannah H Leslie, Stéphane Verguet, et al. Quality of basic maternal care functions in health facilities of five African countries: an analysis of national health system surveys. *The Lancet Global Health*, 4(11):e845–e855, November 2016. Number: 11.

[51] Marie Ng, Nancy Fullman, Joseph L. Dieleman, et al. Effective Coverage: A Metric for Monitoring Universal Health Coverage. *PLOS Medicine*, 11(9):e1001730, September 2014. Publisher: Public Library of Science.

[52] Hannah H. Leslie, Zeye Sun, and Margaret E. Kruk. Association between infrastructure and observed quality of care in 4 healthcare services: A cross-sectional study of 4,300 facilities in 8 countries. *PLoS Med*, 14(12):e1002464, December 2017.

[53] Agbessi Amouzou, Hannah Hogan Leslie, Malathi Ram, et al. Advances in the measurement of coverage for RMNCH and nutrition: from contact to effective coverage. *BMJ Global Health*, 4(Suppl 4):e001297, May 2019. Publisher: BMJ Specialist Journals Section: Analysis.

[54] Sudip Bhandari and Olakunle Alonge. Measuring the resilience of health systems in low- and middle-income countries: a focus on community resilience. *Health Res Policy Sys*, 18(1):81, December 2020. Number: 1.

[55] Yuen W. Hung, Klesta Hoxha, Bridget R. Irwin, et al. Using routine health information data for research in low- and middle-income countries: a systematic review. *BMC Health Services Research*, 20(1):790, August 2020.

[56] H Leslie, D Spiegelman, X Zhou, et al. Service readiness of health facilities in Bangladesh, Haiti, Kenya, Malawi, Namibia, Nepal, Rwanda, Senegal, Uganda and the United Republic of Tanzania. *Bull. World Health Organ.*, 95(11):738–748, November 2017.

[57] Anna D Gage, Hannah H Leslie, Asaf Bitton, et al. Assessing the quality of primary care in Haiti. *Bull. World Health Organ.*, 95(3):182–190, March 2017.

[58] Hannah H. Leslie, Xin Zhou, Donna Spiegelman, et al. Health system measurement: Harnessing machine learning to advance global health. *PLoS One*, 13(10):e0204958, 2018.

[59] Laura Di Giorgio, David K Evans, Magnus Lindelow, et al. Analysis of clinical knowledge, absenteeism and availability of resources for maternal and child health: a cross-sectional quality of care study in 10 African countries. *BMJ Glob Health*, 5(12):e003377, December 2020.

[60] Avedis Donabedian. The Quality of Medical Care. *Science*, 200(4344):856–864, 1978.

[61] World Health Organization. Service Availability and Readiness Assessment (SARA). An annual monitoring system for service delivery. Technical report, World Health Organization, Geneva, Switzerland, 2013.

[62] WHO, editor. *Integrated management of childhood illness: chart booklet.* IMCI, Integrated Management of Childhood Illness. World Health Organization, Geneva, 2014.

[63] Corrina Moucheraud and Kaitlyn McBride. Variability in Health Care Quality Measurement among Studies Using Service Provision Assessment Data from Low- and Middle-Income Countries: A Systematic Review. *The American Journal of Tropical Medicine and Hygiene*, 103(3):986–992, September 2020.

[64] Theodros Getachew, Solomon Mekonnen Abebe, Mezgebu Yitayal, et al. Assessing the quality of care in sick child services at health facilities in Ethiopia. *BMC Health Services Research*, 20(1):574, June 2020.

[65] Cheikh Mbacké Faye, Fernando C. Wehrmeister, Dessalegn Y. Melesse, et al. Large and persistent subnational inequalities in reproductive, maternal, newborn and child health intervention coverage in sub-Saharan Africa. *BMJ Global Health*, 5(1):e002232, January 2020. Publisher: BMJ Specialist Journals Section: Analysis.

[66] Robert C. Reiner, Nicholas Graetz, Daniel C. Casey, et al. Variation in Childhood Diarrheal Morbidity and Mortality in Africa, 2000–2015. *New England Journal of Medicine*, 379(12):1128–1138, September 2018. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa1716766.

[67] Robert C. Reiner, Catherine A. Welgan, Daniel C. Casey, et al. Identifying residual hotspots and mapping lower respiratory infection morbidity and mortality in African children from 2000 to 2017. *Nature Microbiology*, 4(12):2310–2318, December 2019. Number: 12 Publisher: Nature Publishing Group.

[68] Robert C Reiner, Kirsten E Wiens, Aniruddha Deshpande, et al. Mapping geographical inequalities in childhood diarrhoeal morbidity and mortality in low-income and middle-income countries, 2000–17: analysis for the Global Burden of Disease Study 2017. *The Lancet*, 395(10239):1779–1801, June 2020.

[69] Laina D. Mercer, Fred Lu, and Joshua L. Proctor. Sub-national levels and trends in contraceptive prevalence, unmet need, and demand for family planning in Nigeria with survey uncertainty. *BMC Public Health*, 19(1):1752, December 2019.

[70] Nicole A Perales, Dorothy Wei, Aayush Khadka, et al. Quality of clinical assessment and child mortality: a three-country cross-sectional study. *Health Policy and Planning*, 35(7):878–887, August 2020.

[71] Winfred Dotse-Gborgbortsi, Andrew J. Tatem, Victor Alegana, et al. Spatial inequalities in skilled attendance at birth in Ghana: a multilevel

analysis integrating health facility databases with household survey data. *Tropical Medicine & International Health*, 25(9):1044–1054, 2020. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/tmi.13460.

[72] Adrien Allorant, Canada Parrish, Gracia Desforges, et al. Closing the gap in implementation of HIV clinical guidelines in a low resource setting using electronic medical records. *BMC Health Services Research*, 20(1):804, August 2020.

[73] Tarun Gera, Dheeraj Shah, Paul Garner, et al. Integrated management of childhood illness (IMCI) strategy for children under five. *The Cochrane Database of Systematic Reviews*, pages 1–61, July 2016.

[74] Joanna RM Armstrong Schellenberg, Taghreed Adam, and et al. Victora, Cesar. Effectiveness and cost of facility-based integrated management of childhood illness (IMCI) in tanzania. *The Lancet*, 364(9445):1583–1594, 2004. Publisher: Elsevier.

[75] Carsten Krüger, Monika Heinzel-Gutenbrunner, and Mohammed Ali. Adherence to the integrated management of childhood illness guidelines in namibia, kenya, tanzania and uganda: evidence from the national service provision assessment surveys. *BMC Health Services Research*, 17(1):822, 2017.

[76] Laina D. Mercer, Jon Wakefield, Athena Pantazis, et al. Space–time smoothing of complex survey data: Small area estimation for child mortality. *Ann. Appl. Stat.*, 9(4):1889–1905, December 2015.

[77] Andrea Riebler, Sigrunn H. Sørbye, Daniel Simpson, et al. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *arXiv:1601.01180 [stat]*, January 2016. arXiv: 1601.01180.

[78] D. J. Weiss, A. Nelson, H. S. Gibson, et al. A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, 553(7688):333–336, January 2018. Number: 7688 Publisher: Nature Publishing Group.

[79] Nicholas Graetz, Joseph Friedman, Aaron Osgood-Zimmerman, et al. Mapping local variation in educational attainment across Africa. *Nature*, 555(7694):48–53, March 2018. Number: 7694 Publisher: Nature Publishing Group.

[80] Matti Kummu, Maija Taka, and Joseph H. A. Guillaume. Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015. *Sci Data*, 5(1):180004, December 2018.

[81] Miguel A. Martinez-Beneito and Paloma Botella-Rocamora. *Disease Mapping : From Foundations to Multidimensional Modeling.* Chapman and Hall/CRC, July 2019.

[82] J. N. K. Rao and Isabel Molina. *Small area estimation.* Wiley series in survey methodology. John Wiley & Sons, Inc, Hoboken, New Jersey, second edition edition, 2015.

[83] USAID. USAID. Senegal Health Project 2016-2021. Technical report, USAID, 2015.

[84] Cesar Gomes Victora, Gary Joseph, Inacio C.M. Silva, et al. The Inverse Equity Hypothesis: Analyses of Institutional Deliveries in 286 National Surveys. *Am J Public Health*, 108(4):464–471, February 2018. Publisher: American Public Health Association.

[85] Matthew Barclay, Mary Dixon-Woods, and Georgios Lyratzopoulos. The problem with composite indicators. *BMJ Qual Saf*, 28(4):338–344, April 2019. Publisher: BMJ Publishing Group Ltd Section: The problem with. . . .

[86] Ambrose Agweyu, Theopista Masenge, and Deogratias Munube. Extending the measurement of quality beyond service delivery indicators. *BMJ Global Health*, 5(12):e004553, December 2020. Publisher: BMJ Specialist Journals Section: Editorial.

[87] Siri Lange, Aziza Mwisongo, and Ottar Mæstad. Why don't clinicians adhere more consistently to guidelines for the Integrated Management of Childhood Illness (IMCI)? *Social Science & Medicine*, 104:56–63, March 2014.

[88] May Sudhinaraset, Matthew Ingram, Heather Kinlaw Lofthouse, and Dominic Montagu. What Is the Role of Informal Healthcare Providers in Developing Countries? A Systematic Review. *PLOS ONE*, 8(2):e54978, 2013. Publisher: Public Library of Science.

[89] Papa Yona Boubacar Mané. Efficience et équité dans le système de santé du Sénégal. page 281.

[90] Francesca L. Cavallaro, Lenka Benova, El Hadji Dioukhane, et al. What the percentage of births in facilities does not measure: readiness for emergency obstetric care and referral in Senegal. *BMJ Global Health*, 5(3):e001915, March 2020. Publisher: BMJ Specialist Journals Section: Original research.

[91] Ariel BenYishay, Daniel Runfola, Rachel Trichler, et al. A Primer on Geospatial Impact Evaluation Methods, Tools, and Applications. AidData Working Paper #44., 2017.

[92] Gavi Full Country Evaluations team. Gavi Full Country Evaluations. 2016 Annual Dissemination Report. Cross-Country Findings. Technical report, GAVI, 2017.

[93] Martha Priedeman Skiles, Clara R Burgert, Siân L Curtis, et al. Geographically linking population and facility surveys: methodological considerations. *Popul Health Metrics*, 11(1):14, December 2013.

[94] Ryan A. Simmons, Rebecca Anthopolos, and Wendy Prudhomme O'Meara. Effect of health systems context on infant and child mortality in sub-Saharan Africa from 1995 to 2015, a longitudinal cohort analysis. *Sci Rep*, 11(1):16263, December 2021.

[95] Katherine R. Paulson, Aruna M. Kamath, Tahiya Alam, et al. Global, regional, and national progress towards Sustainable Development Goal 3.2 for neonatal and child health: all-cause and cause-specific mortality findings from the Global Burden of Disease Study 2019. *The Lancet*, 398(10303):870–905, September 2021. Publisher: Elsevier.

[96] Ties Boerma, Patrick Eozenou, David Evans, et al. Monitoring Progress towards Universal Health Coverage at Country and Global Levels. *PLOS Medicine*, 11(9):e1001731, September 2014. Publisher: Public Library of Science.

[97] Adam Wagstaff and Sven Neelsen. A comprehensive assessment of universal health coverage in 111 countries: a retrospective observational study. *The Lancet Global Health*, 8(1):e39–e49, January 2020.

[98] T. Tanahashi. Health service coverage and its evaluation. *Bull World Health Organ*, 56(2):295–303, 1978.

[99] Peter K. Nguhiu, Edwine W. Barasa, and Jane Chuma. Determining the effective coverage of maternal and child health services in Kenya, using demographic and health survey data sets: tracking progress towards universal health coverage. *Trop Med Int Health*, 22(4):442–453, April 2017.

[100] Hannah H Leslie, Address Malata, Youssoupha Ndiaye, et al. Effective coverage of primary care services in eight high-mortality countries. *BMJ Glob Health*, 2(3):e000424, September 2017.

[101] Elysia Larson, Daniel Vail, Godfrey M. Mbaruku, et al. Beyond utilization: measuring effective coverage of obstetric care along the quality cascade. *Int J Qual Health Care*, 29(1):104–110, February 2017.

[102] Agbessi Amouzou, Safia S. Jiwani, Inácio Crochemore Mohnsam da Silva, et al. Closing the inequality gaps in reproductive, maternal, newborn and child health coverage: slow and fast progressors. *BMJ Global Health*, 5(1):e002230, January 2020. Publisher: BMJ Specialist Journals Section: Original research.

[103] Andrew D. Marsh, Moise Muzigaba, Theresa Diaz, et al. Effective coverage measurement in maternal, newborn, child, and adolescent health and nutrition: progress, future prospects, and implications for quality health systems. *The Lancet Global Health*, 8(5):e730–e736, May 2020. Publisher: Elsevier.

[104] Jonathan F Mosser, William Gagne-Maynard, Puja C Rao, et al. Mapping diphtheria-pertussis-tetanus vaccine coverage in Africa, 2000–2016: a spatial and temporal modelling study. *The Lancet*, 393(10183):1843–1855, May 2019.

[105] Alyssa N. Sbarra, Sam Rolfe, Jason Q. Nguyen, et al. Mapping routine measles vaccination in low- and middle-income countries. *Nature*, 589(7842):415–419, January 2021. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 7842 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Disease prevention;Infectious diseases;Public health Subject_term_id: disease-prevention;infectious-diseases;public-health.

[106] Aniruddha Deshpande, Molly K. Miller-Petrie, Paulina A. Lindstedt, et al. Mapping geographical inequalities in access to drinking water and sanitation facilities in low-income and middle-income countries, 2000–17. *The Lancet Global Health*, 8(9):e1162–e1185, September 2020. Publisher: Elsevier.

[107] Riese Sara, Shireen Assaf, and Thomas Pullum. Measurement approaches for effective coverage estimation. DHS methodological reports no. 31. Technical report, Demographic and Health Survey, 2021.

[108] Deepa Jahagirdar, Magdalene K. Walters, Amanda Novotney, et al. Global, regional, and national sex-specific burden and control of the HIV epidemic, 1990–2019, for 204 countries and territories: the Global Burden of Diseases Study 2019. *The Lancet HIV*, 8(10):e633–e651, October 2021. Publisher: Elsevier.

[109] Amelia Bertozzi-Villa, Caitlin A. Bever, Hannah Koenker, et al. Maps and metrics of insecticide-treated net access, use, and nets-per-capita in Africa from 2000-2020. *Nat Commun*, 12(1):3589, June 2021. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Computational models;Epidemiology;Lifestyle modification;Malaria Subject_term_id: computational-models;epidemiology;lifestyle-modification;malaria.

[110] Philippe Apparicio, Mohamed Abdelmajid, Mylène Riva, et al. Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues. *International Journal of Health Geographics*, 7(1):7, February 2008.

[111] Kirsten E. Wiens, Lauren E. Schaeffer, Samba O. Sow, et al. Oral rehydration therapies in Senegal, Mali, and Sierra Leone: a spatial analysis of changes over time and implications for policy. *BMC Med*, 18(1):405, December 2020.

[112] L. Bernadinelli, C. Pascutto, N. G. Best, et al. DISEASE MAPPING WITH ERRORS IN COVARIATES. *Statist. Med.*, 16(7):741–752, April 1997.

[113] Stefanie Muff, Andrea Riebler, Leonhard Held, et al. Bayesian analysis of measurement error models using integrated nested Laplace approximations. *J. R. Stat. Soc. C*, 64(2):231–252, February 2015.

[114] Merlin L Willcox, Elias Kumbakumba, Drissa Diallo, et al. Circumstances of child deaths in Mali and Uganda: a community-based confidential enquiry. *The Lancet Global Health*, 6(6):e691–e702, June 2018.

[115] Christy Hanson, Mike Osberg, Jessie Brown, et al. Finding the Missing Patients With Tuberculosis: Lessons Learned From Patient-Pathway Analyses in 5 Countries. *The Journal of Infectious Diseases*, 216(suppl_7):S686–S695, November 2017.

[116] Katherine C. Horton, Peter MacPherson, Rein M. G. J. Houben, et al. Sex Differences in Tuberculosis Burden and Notifications in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis. *PLoS Med*, 13(9):e1002119, September 2016.

[117] Ahmad Reza Hosseinpoor, Nicole Bergen, Aluisio J. D. Barros, et al. Monitoring subnational regional inequalities in health: measurement approaches and challenges. *Int J Equity Health*, 15(1):18, December 2016.

[118] Michele Nguyen, Rosalind E. Howes, Tim C.D. Lucas, et al. Mapping malaria seasonality in Madagascar using health facility data. *BMC Medicine*, 18(1):26, February 2020.

Appendix A

APPENDIX: CHAPTER 2

A.1 Supplementary Data

A.2 Supplementary Methods

A.2.1 Estimating tuberculosis prevalence at the district level

District-level TB prevalence using Horvitz-Thompson estimator

For each cluster $c = 1, \ldots, 125$ sampled within strata (rural/urban), $k \in s_c$ represents the individuals sampled in cluster c. We label $Y_{ck} = 1$ if individual k in cluster c is a confirmed TB case. Noting design-weights w_{ck}, we estimate the prevalence of TB in district i using the design-based weighted or Horvitz-Thompson estimator:

$$p_i^{HT} = \frac{\sum_{c \in i} \sum_{k \in c} w_{ck} Y_{ck}}{\sum_{c \in i} \sum_{k \in c} w_{ck}} \tag{A.1}$$

with variance V_i^* calculated using standard methods, such as jackknife. Supplemental Figure S1 displays the design-based estimates of prevalence and their associated uncertainty for the 45 districts where data was collected.

Spatial smoothing model

To increase the precision of the design-based estimates of prevalence, and to predict prevalence for districts where no data was collected, we use a hierarchical spatial model inspired by Fay and Herriot classic framework:

$$\text{logit}(p_i^{HT} \sim N(\theta_i, V_i) \tag{A.2}$$

$$\theta_i = x_i^T \beta + \varepsilon_i + S_i \tag{A.3}$$

Table A.1: Covariate data sources

Covariate	Spatial resolution	Temporal resolution	Source
Outdoor air pollution[a]	Pixel-level	Annual	IHME
Household air pollution[a]	District	Annual	IHME
Average daily mean temperature	Pixel-level	Annual	CRUTS
Population	Pixel-level	Annual	WorldPop
Travel time to nearest settlement >50,000 inhabitants[a]	Pixel-level	Annual	WorldPop
Urban proportions by cluster	Pixel-level	2015	ESA land cover
Population density	Pixel-level	Annual	WorldPop
Coverage BCG Vaccine	Pixel-level	Annual	IHME
Nighttime lights[a]	Pixel-level	Annual	VIIRS
Mean wealth index[a]	Pixel-level	2011	WorldPop
Protein daily[a]	District	2016	World Bank
Educational attainment	District	2016	World Bank
% population less than 15, between 15-64, above 65[a]	District	2016	World Bank
% population with primary employment in agriculture	District	2016	World Bank
Poverty headcount ratio	District	2016	World Bank
% population in extreme poverty	District	2016	World Bank
Severely stunted (rank)[a]	District	2016	World Bank
Household crowding[a]	District	2007-14	DHS (derived)

[a]Selected with the variance inflation factor (with a threshold of 5)

Where $\varepsilon_i \sim_{iid} N(0, \sigma_\varepsilon^2)$ and $S = [S_1, \ldots, S_n]$ are intrinsic conditional autoregressive (ICAR) terms. Our estimate of prevalence in district i is then $\widehat{p_i} = e^{\widehat{\varepsilon_i}}$

Model selection

Our model selection criteria favored parsimonious models with fewer predictors: the best model according to the DIC, WAIC and RMSE only included the temperature, while the CPO favored a model with both temperature and prevalence of overcrowding as predictors. Both models, along with a model without any predictor (simple spatial smoothing), produced similar predicted TB prevalence rates.

A.2.2 Estimating prevalence-to-notification ratio at the district level

For each district i we calculated for 2016 the case notification rate defined as the number of TB case notifications per 100,000 inhabitants, $cnr_i = \frac{N_i}{pop_i} * 100,000$, where N_i is the number of cases notified in district i in 2016, and pop_i is the population in district i. We then obtained our estimate of the prevalence-to-notification ratio (with uncertainty interval) as the ratio of the prevalence estimate and the case notification rate: $pn_i = \frac{\hat{p_i}}{cnr_i}$

A.2.3 Counterfactual analysis

For each district, we calculated the additional number of cases that could have been notified had the district at least achieved the national prevalence to notification ratio of 2.8 (noted N_i^*).

The counterfactual number of cases notified in district i noted N_i^c is equal to:

$$N_i^c = \frac{\hat{p_i} * 100,000}{\min(pn_i, 2.8)} \tag{A.4}$$

We then derive N_i^* from the counterfactual number of cases notified in district i and the actual number of cases notified in district i, N_i:

$$N_i^* = N_i^c - N_i \tag{A.5}$$

By definition, N_i^* is zero in all districts were the estimated prevalence-to-notification ratio is less or equal to 2.8. The case notification rate is defined as the number of TB case notifications per 100,000 inhabitants, and noted cnr_i. We then obtained our estimate of the prevalence-to-notification ratio (with uncertainty interval) as the ratio of the prevalence estimate and the case notification rate: $pn_i = \frac{\hat{p_i}}{cnr_i}$

A.3 Supplementary Results

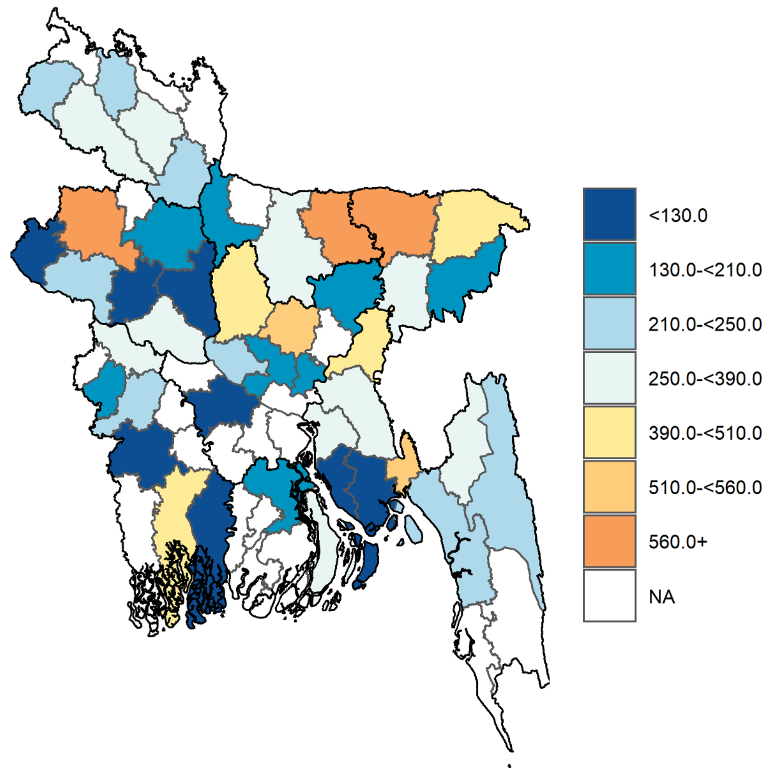


Figure A.1: Design-based (Horvitz-Thompson) estimates of TB prevalence by district, 2015–2016 Districts left blank (labelled "NA") are the districts that did not contain any cluster sampled during the survey in 2015–2016.

Table A.2: Model comparison using different set of covariates)

Model Number	Model (covariates number)	DIC	p_{eff}	$\sum \log \text{CPO}$	WAIC	RMSE
1	None (smoothing model)	33.83355	19.3426	-26.989	32.19276	9.68E-04
2	1,3,9,10,13,17,18	35.64441	19.52178	-29.7489	34.77869	9.77E-04
3	1,2,3,5,9,10,11,13,17,18	35.65755	21.99758	-31.9772	32.88017	9.59E-04
4	1,3,9,10,13,18	35.29093	17.04275	-28.107	35.58705	0.001053
5	1,3,10,13,17,18	33.52514	17.92275	-26.377	32.68829	9.81E-04
6	2,3,9,18	31.56231	17.60470	-25.8767	29.78304	9.56E-04
7	1,3,10,1	32.6375	17.4566	-26.9979	31.559	9.40E-04
8	3,18[b]	29.2925	16.78769	-23.0158	27.25019	9.03E-04
9	3,17,18	30.1587	17.50217	-23.8731	27.93195	8.90E-04
10	1,2	34.71324	19.13242	-26.8717	33.1516	9.56E-04
11	18	33.14513	18.04545	-25.3774	32.07906	0.001042
12	2,3	29.98399	17.75053	-23.9708	27.23883	8.56E-04
13	3[a]	28.93787	17.33569	-23.4107	26.23411	8.35E-04
14	2	34.76387	19.55197	-27.405	33.17001	9.99E-04

[a]Best model according to DIC, WAIC and RMSE (model used for maps in manuscript)

[b]Best model according to $\sum \log \text{CPO}$

Table A.3: Predictors' effect estimates

Covariate	Best Model (13)	Best Model CPO (8)	Model 6	Model 7
(Intercept)		30.8 [5.0;58.0]	18.3 [-19.9;59.0]	17.14 [-22.97;60.61]
log(Ambient air temperature)	-8.0 [-12.8;-3.3]	-6.8 [-11.6;-2.1]	-4.5 [-11.8;2.43]	-4.18 [-12.14;3.17]
Household overcrowding		1.8 [-0.4;4.0]	2.2 [-0.15;4.4]	2.40 [-0.12;4.78]
log(Nighttime lights)			-0.017 [-0.25; 0.22]	
Outdoor air pollution				-.006 [-.018; .007]
Household air pollution				0.35 [-.35;1.02]
Precision spatial random effects	1844.9 (1831.0)	1788.3 (1799.2)	1812.5 (1812.76)	

[a]Covariates are estimated at the level of the district from sources detailed in Table A.1

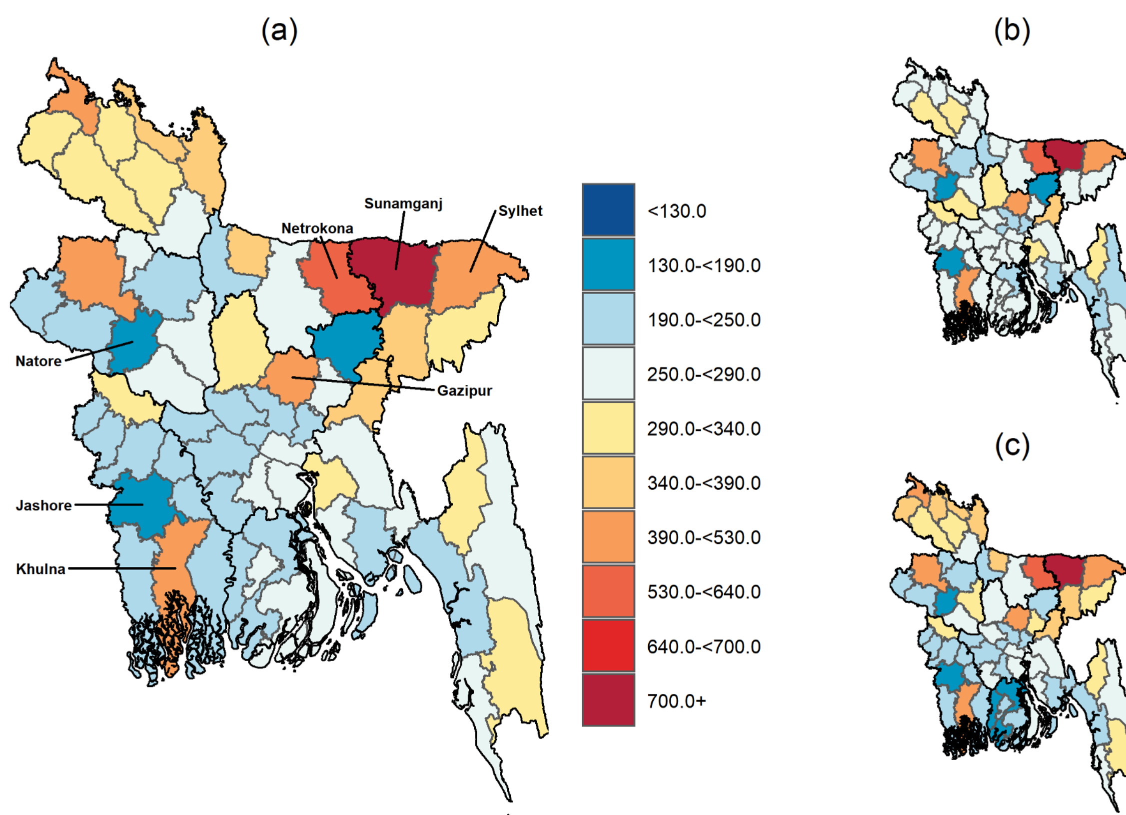


Figure A.2: Predicted TB median prevalence (per 100,000) with the best (a), the smoothing (b) and the CPO (c) model

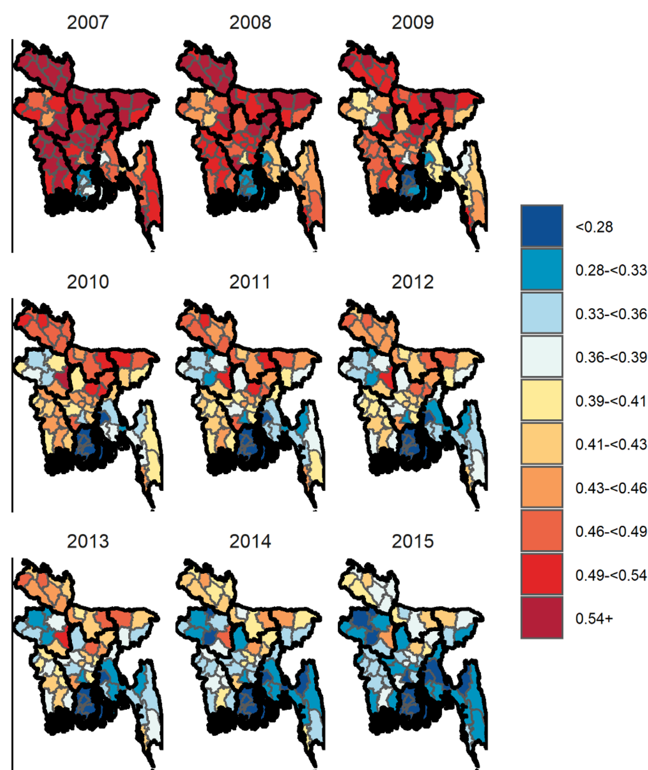


Figure A.3: Predicted median prevalence of household overcrowding in Bangladesh between 2007 and 2015

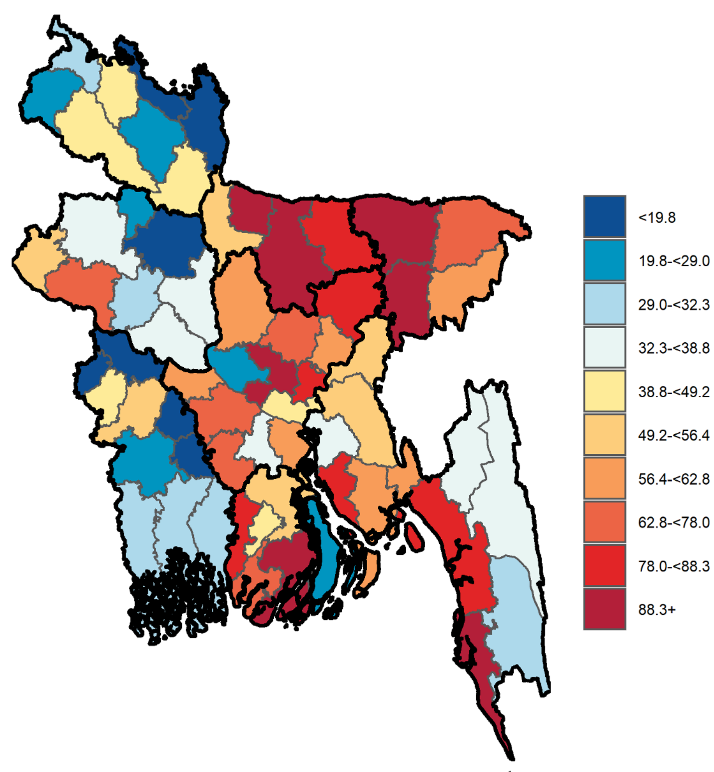


Figure A.4: Relapse cases per 1,000 TB cases in Bangladesh in 2016

Table A.4: Complete list of district-level prevalence estimates per 100,000 inhabitants (1/3)

Districts	Mean estimates	97.5th percentile	2.5th percentile
Barisal	200	280	140
Bhola	260	260	250
Pirojpur	230	490	110
Brahamanbaria	350	580	220
Chandpur	290	360	240
Chittagong	240	370	160
Comilla	270	450	160
Feni	280	580	130
Khagrachhari	310	410	230
Lakshmipur	270	560	130
Noakhali	220	390	110
Rangamati	250	330	190
Dhaka	200	300	130
Faridpur	240	510	120
Gazipur	440	640	310
Gopalganj	240	490	110
Kishoreganj	190	270	130
Manikganj	250	420	150
Narayanganj	200	290	140
Tangail	290	560	150
Bagerhat	220	460	100
Chuadanga	210	400	110
Jessore	160	310	80
Jhenaidah	230	370	140

Table A.5: Complete list of district-level prevalence estimates per 100,000 inhabitants (2/3)

Districts	Mean estimates	97.5th percentile	2.5th percentile
Khulna	400	480	340
Kushtia	310	410	230
Jamalpur	220	330	150
Mymensingh	270	380	190
Netrakona	640	690	580
Bogra	210	320	140
Naogaon	430	650	290
Natore	180	340	90
Nawabganj	230	480	110
Pabna	270	490	150
Rajshahi	210	280	160
Sirajganj	250	530	120
Dinajpur	310	550	180
Gaibandha	260	430	160
Nilphamari	320	670	150
Rangpur	310	470	210
Thakurgaon	320	670	150
Habiganj	340	720	160
Maulvibazar	340	630	170
Sunamganj	840	1020	690
Sylhet	520	710	370
Bandarban	300	630	140
Barguna	240	500	110
Cox's Bazar	260	540	120

Table A.6: Complete list of district-level prevalence estimates per 100,000 inhabitants (3/3)

Districts	Mean estimates	97.5th percentile	2.5th percentile
Jhalokati	250	530	120
Joypurhat	270	560	130
Kurigram	350	740	170
Lalmonirhat	340	710	160
Madaripur	250	520	120
Magura	230	490	110
Meherpur	210	450	100
Munshiganj	260	530	120
Narail	220	470	110
Narsingdi	280	570	130
Panchagarh	410	860	190
Patuakhali	260	540	120
Rajbari	240	500	110
Satkhira	200	420	90
Shariatpur	260	540	120
Sherpur	370	770	170

Appendix B

APPENDIX: CHAPTER 3

B.1 Supplementary data

B.1.1 Samples of the SDI and SPA surveys

In Senegal, data were collected as part of a **Continuous SPA**, which achieved a census of all facilities after five rounds of surveys. Half of the hospitals and health centers were selected in the first (2012-13) and third round (2015) of SPA, while the other half were selected in the second (2014) and fourth round (2016), which created a dependent sampling structure between the first four rounds of the continuous SPA. Random samples of facilities were selected for the SPA conducted in 2017, 2018 and 2019. As the higher levels in the process quality metrics, observed in years 2015 and 2016 compared to 2012-13 and 2014, were not found in the later years, we can hypothesize that the facilities and providers who were observed for the second time in 2015 and 2016, had a better knowledge of SPA interviewers assessment criteria, which could have led to an enhanced Hawthorne effect.

B.1.2 Geo-locating health facilities

For SPA survey data, we determined in which regions each facility was located using the GPS coordinates provided by the DHS program and the second-level administrative shapefiles that are publicly available from the Database of Global Administrative Areas (https://gadm.org/). The two most recent SPA surveys in Senegal did not collect GPS coordinates, but DHS provided a linkage file to assign a *departement* to each sampled facility. SDI surveys include the administrative units as a variable directly.

Table B.1: Characteristics of the health facilities sampled in the SPA, by year in Senegal, Kenya and Tanzania

	Senegal							Kenya			Tanzania	
	SPA 2013	SPA 2014	SPA 2015	SPA 2016	SPA 2017	SPA 2018	SPA 2019	SPA 1999	SPA 2004	SPA 2010	SPA 2006	SPA 2014
Total number of facilities	364	363	375	371	396	339	361	388	440	703	611	1200
Facility type												
Hospital	35	35	38	35	35	29	31	32	172	253	128	263
Health centre	64	62	65	61	74	62	64	90	51	101	41	380
Clinics	265	266	272	275	287	248	246	256	217	349	437	557
Managing authority												
Public	305	293	299	288	315	270	288	177	175	351	425	783
Private	59	70	76	83	81	69	73	211	265	352	186	417
Sick child consultations observed	1307	1212	1262	1026	1064	715	884	623	1211	2016	2559	4961

B.1.3 Sampling of sick child visits and vignettes

In the SPA, the survey teams randomly selected three providers of curative care of sick children among all providers in this service present at the facility the day of the assessment. A maximum of five client consultations for each selected provider was observed. The client weights use the facility sampling weight as its base weight, and take into account the total number of clients listed and interviewed within each of the sampling stratum, to calculate the probability of a given consultation to be observed. In the SDI, vignettes were administered to a random sample of health providers (doctors, medical assistants, nurses) among all eligible providers present at the facility the day of the assessment. providers were randomly sampled among all eligible providers at the facility. The weights here simply represents the percentage probability of selection of a provider within each facility.

B.1.4 Differences between the SPA and SDI surveys

The SPA and SDI surveys' inventory questionnaire mostly collect the same information, as both surveys use WHO's SARA framework to assess the availability and readiness of key infrastructure and services in facilities, what we referred to as readiness in this study. However, the differences in the methodology used to assess providers' knowledge and competence, referred to as process quality in this article, differs significantly. In the SPA survey, interviewers observe directly patient-provider consultations using an observation protocol. In the SDI survey, interviewers act as patients and record providers' questions, examinations and recommendations to an hypothetical clinical case, known as vignette. Past studies have shown that compliance to protocol checklists tend to be higher with vignettes than with direct clinical observations [59]. To account for these differences in the assessment of process quality between the SPA and SDI surveys, we used survey-specific effects in our model.

B.2 Supplementary Methods

B.2.1 Spatial random effects

Regional-level random effects were used to model the potential influence of unmeasured features of the local health system. The use of spatially structured random effects reflects our assumption that these unobserved characteristics affecting the availability of quality care are likely to be correlated in space. For instance, remote administrative units might face challenges - in accessing steady drug supply chains for essential medicines and testing materials, or in attracting qualified health workers - that are more similar to adjacent units than that experienced in the capital city. We evaluated alternative models where regional-level random effects were assumed to be independent and identically distributed rather than spatially correlated (see C.2.1).

B.2.2 Senegal - Accounting for the sampling design

In Senegal, to account for the changing sampling methods (described in B.1.1), we adopted an analytical approach that includes random-effects to account for the rounds of SPA that comprised repeated selection of facilities.

B.2.3 Model selection

Direct estimates and design-based variance estimates were computed using the ***survey*** package in *R*. We fit the Bayesian hierarchical models using the Integrated Nested Laplace Approximation [32] and the R-INLA package [33]. We obtained a subset of all included covariates by checking for multi-collinearity using the variance inflation factor (VIF) with a threshold of 5. For each indicator and country, we compared the 8 models presented in table C.2 consisting of different combinations of the covariates selected with the VIF procedure, and spatio-temporal random effects, using three selection procedures (the deviance information criteria [DIC], the Watanabe-Akaike information criteria [WAIC], and the conditional predictive ordinate [CPO]). In the absence of consensus on a single criterion, when different criteria pointed to different models, we used a majority rule.

B.2.4 Summary inequality measures description

Documenting systematic differences in readiness and process quality can have important implications for policy and drive spatially targeted interventions [117]; we used several absolute and relative metrics of inequality capturing different patterns of disparities.

High-to-low difference and ratio

Comparing posterior samples of quality of care metrics between administrative units, we estimated the high-to-low difference (respectively, ratio) as the absolute difference (resp. relative ratio) between the best (noted p_{nt}) and worst (p_{0t}) performing area in a given country

for a given year:

$$\widehat{p_{nt}}^s - \widehat{p_{0t}}^s \quad \text{for } s = 1, \ldots, 1000$$
$$\frac{\widehat{p_{nt}}^s}{\widehat{p_{0t}}^s} \quad \text{for } s = 1, \ldots, 1000$$

Inequality pattern index

The inequality pattern index proposed by Victora and colleagues [84] is a measure of regional inequality, which describes the difference between two gaps: the gap between the top-performing subnational unit and the national mean, and between the worst-performing subnational unit and the national mean, such that positive values suggest greater gaps between the top performing area compared to the national level, while negative values indicate larger differences between the bottom and the national level:

$$(\widehat{p_{nt}}^s - \widehat{p_t}^s) - (\widehat{p_t}^s - \widehat{p_{0t}}^s) \quad \text{for } s = 1, \ldots, 1000$$

Supplementary Results

B.2.5 Model validation

Figure B.1 displays the posterior uncertainty intervals estimated from the model, and the observed survey estimates for different regions and years in Senegal, for the four readiness and process quality metrics. Figures B.2 and B.3 present the same information for Kenya, and Tanzania, respectively.

Figure B.7 proposes bi-variate maps describing how service contact and receipt of intervention coverage vary together spatially in Senegal, Kenya and Tanzania, over the study period. Administrative units are grouped by quantiles to compare their relative performances on these two dimensions of process quality and identify priority areas. Figure C.1**a-b** suggests that the relative ranking of subnational units has changed substantially over the study

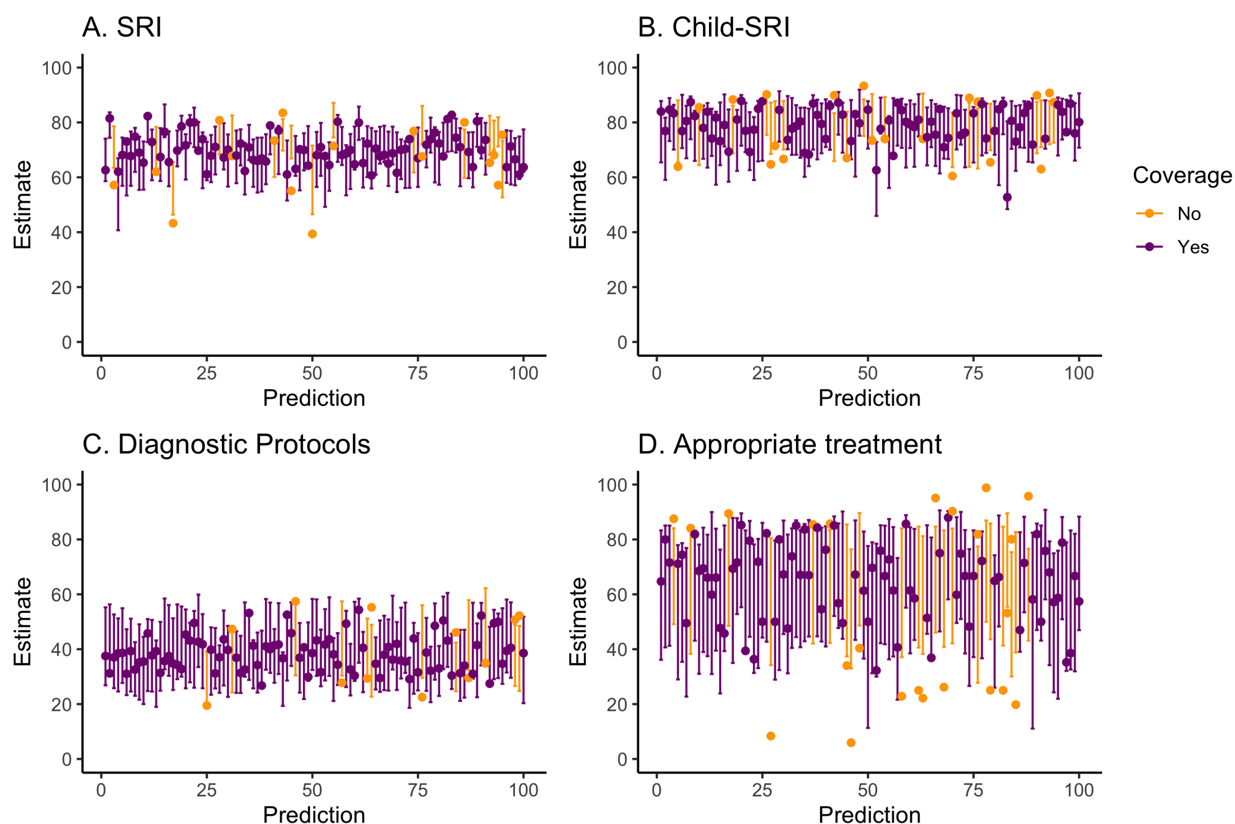


Figure B.1: Direct survey estimates coverage by model-based uncertainty interval, in Senegal.

period. In Kenya, most subnational units in 2020 achieve high service and intervention coverage in the upper right quadrant of figure C.1**g**, i.e above 60%, which suggests important improvements since 2010, when receipt of intervention coverage was below 50% for most subnational units. These changes seem to have enhanced the spatial disparities observed on figure C.1**c-d**: highest coverage of both service contact and receipt of intervention in the center of the country, lower service coverage in the Northeastern part of the country, and lower intervention coverage in the lake region and the southeastern part of the country. In Tanzania, all subnational units are comprised on the upper and lower right quadrant of figure C.1**g**, suggesting levels of service contact above 50%; receipt of intervention coverage is however less than 50% in most subnational units, over the study period, and for all sub-

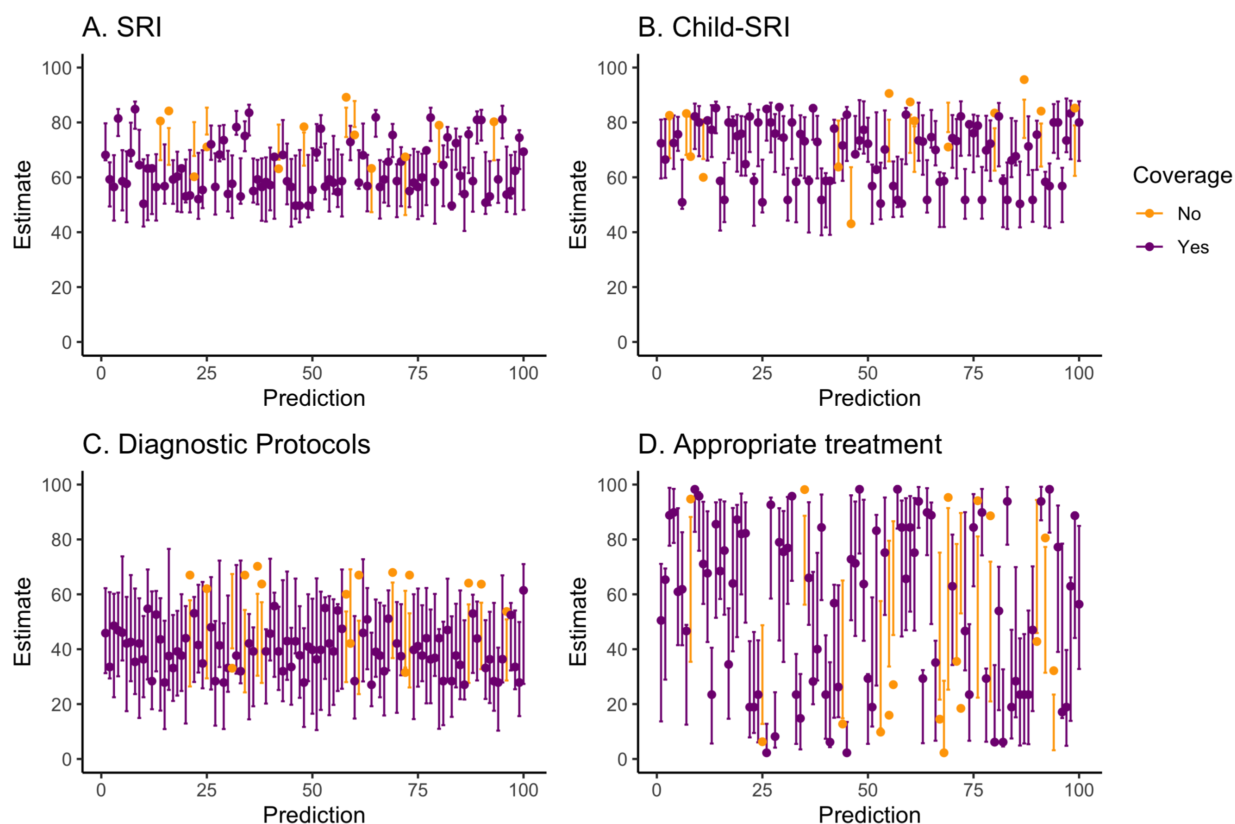


Figure B.2: Direct survey estimates contained in the model-based 95% credible intervals, in Kenya.

national units in 2020. Figure C.1**e-f** suggests consistent spatial patterns between 2010 and 2020, with higher receipt of intervention coverage but lower service coverage in the north-western regions of the country, higher service coverage but lower intervention coverage in the south, and higher levels of both service and intervention coverage in the eastern regions of the country. We further explored subnational disparities in process quality metrics by looking at the relative ranking of subnational units along the two dimensions of providers' average compliance with diagnostic protocols and adequate treatment of sick children (see supplementary figure B.7).

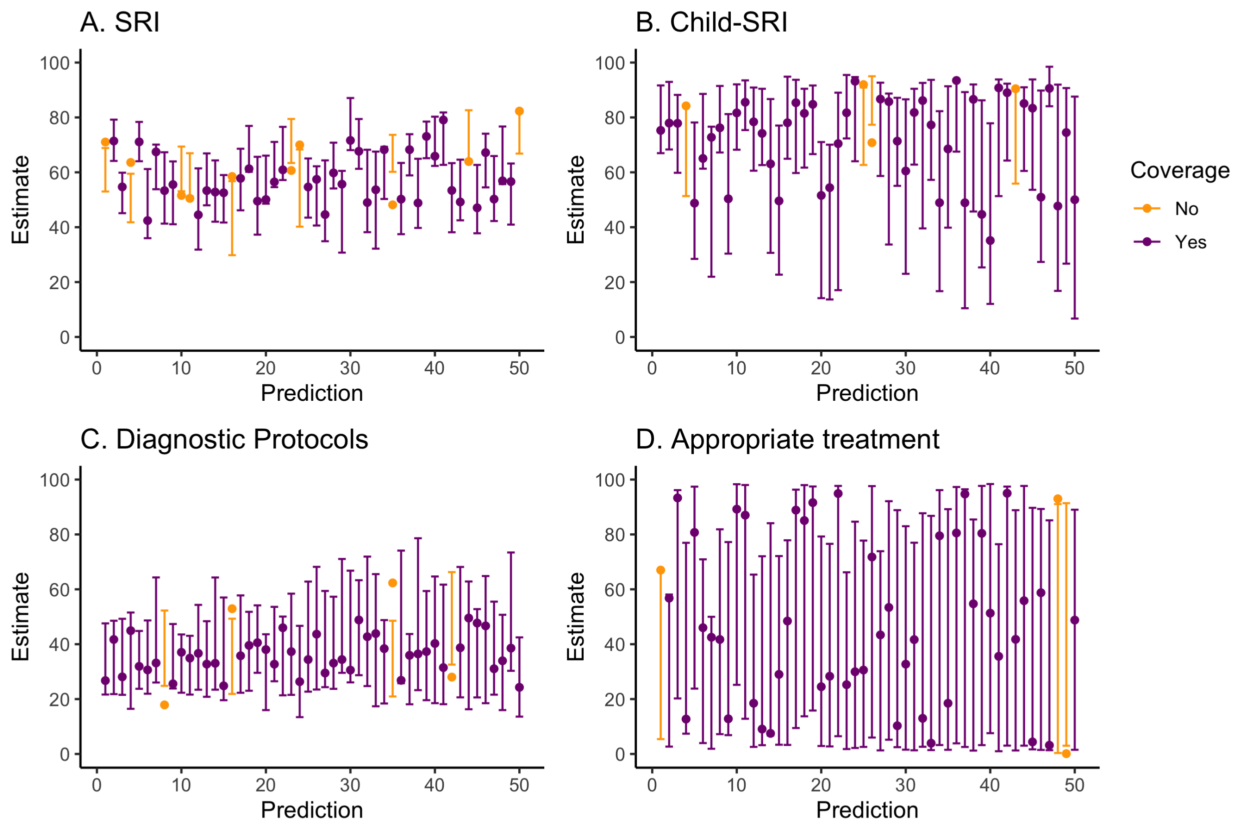


Figure B.3: Direct survey estimates contained in the model-based 95% credible intervals, in Tanzania.

Table B.2: Characteristics of the health facilities sampled in the SDI, by year in Senegal, Kenya and Tanzania

	Senegal	Kenya		Tanzania		
	SDI 2010	SDI 2012	SDI 2018	SDI 2010	SDI 2014	SDI 2016
Total number of facilities	151	292	3,094	175	383	386
Facility type						
Hospital	0	51	161	NA	27	30
Health centre	142	62	484	NA	84	92
Clinics	41	100	2,449	NA	272	264
Managing authority						
Public	151	158	1,781	175	269	273
Private	0	134	1,313	0	134	127
Vignettes	150	625	4,430	180	570	543

Table B.3: Models considered to estimate quality metrics over time and space

Model	Formula
1	$\gamma_t + \alpha_t + e_i + \delta_{it}$
2	$\gamma_t + \alpha_t + e_i + \delta_{it} + \nu_s$
3	$X_{it}\boldsymbol{\beta} + \gamma_t + \alpha_t + e_i + \delta_{it}$
4	$X_{it}\boldsymbol{\beta} + \gamma_t + \alpha_t + e_i + \delta_{it} + \nu_s$
5	$\gamma_t + \alpha_t + e_i + S_i + \delta_{it}$
6	$\gamma_t + \alpha_t + e_i + S_i + \delta_{it} + \nu_s$
7	$X_{it}\boldsymbol{\beta} + \gamma_t + \alpha_t + e_i + S_i + \delta_{it} + \nu_s$

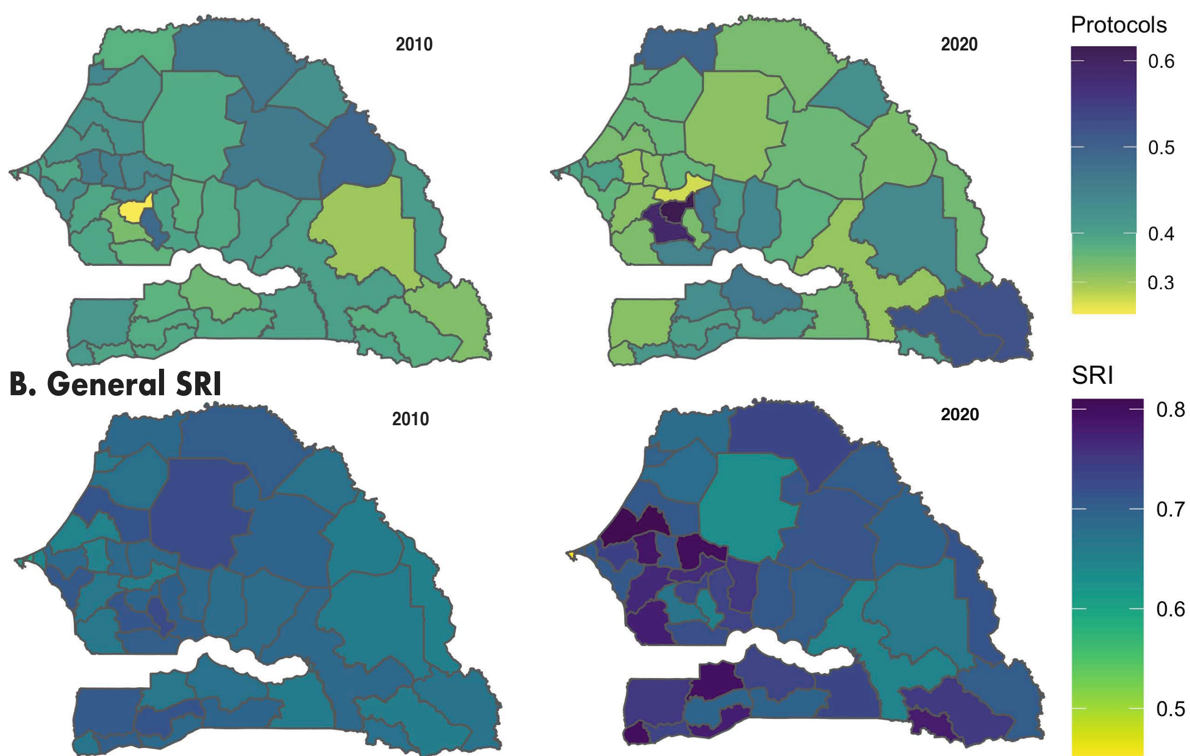
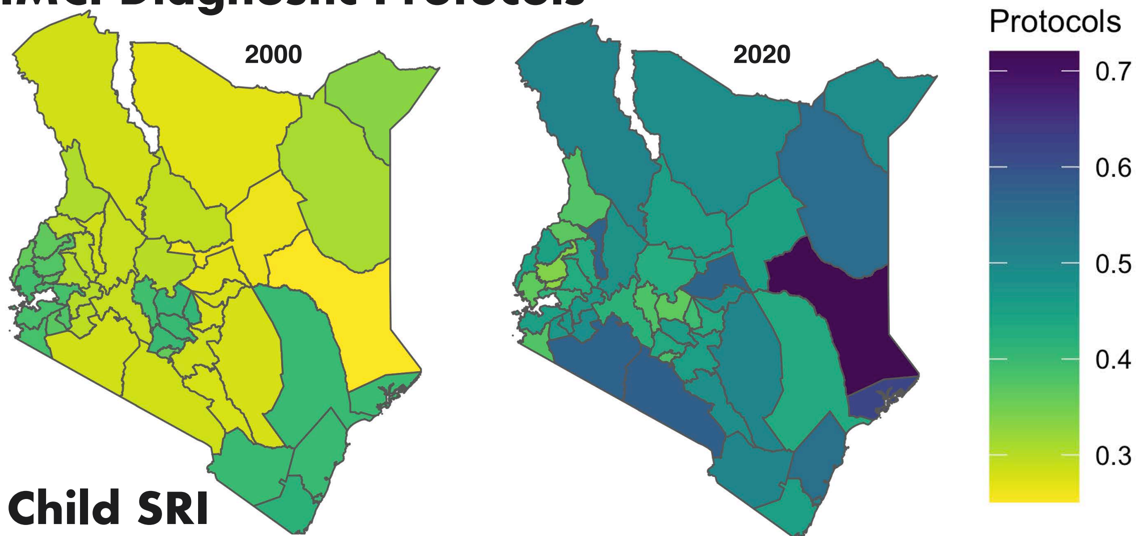


Figure B.4: IMCI diagnostic protocols (a) and General SRI (b) estimates, in Senegal in 2010 and 2020



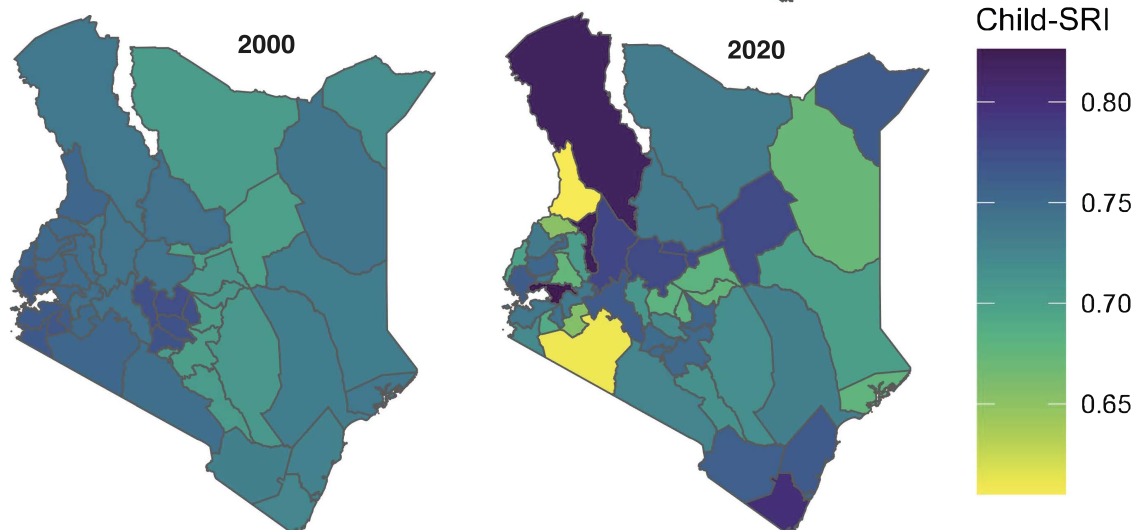


Figure B.5: Child IMCI diagnostic protocols (a) and child SRI (b) estimates, in Kenya in 2000 and 2020

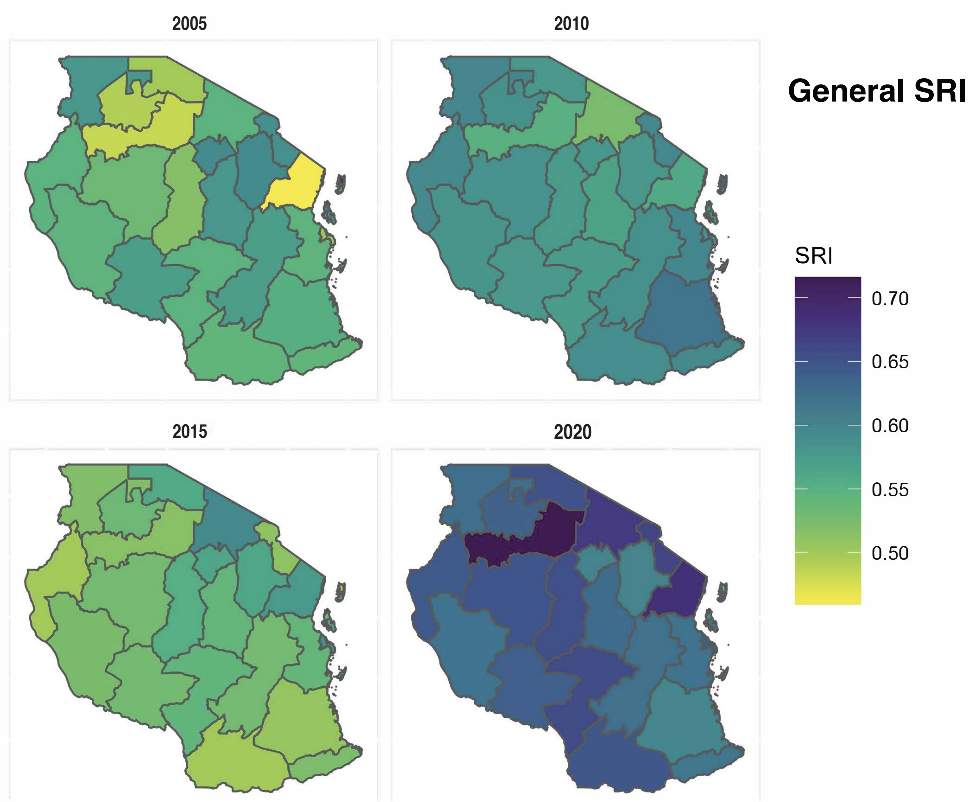


Figure B.6: General SRI estimates, in Tanzania between 2005 and 2020

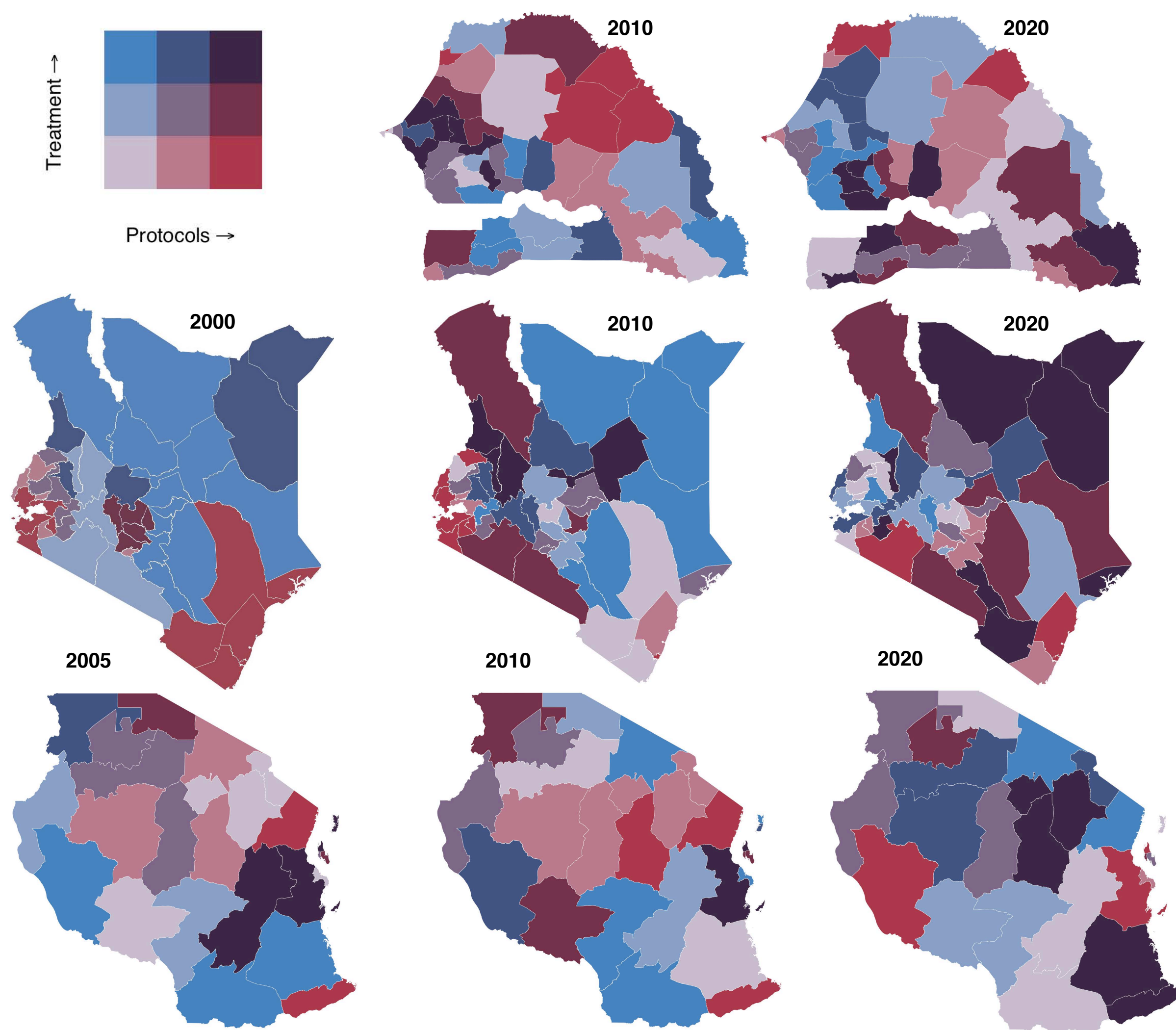


Figure B.7: Joint distribution of the two process quality of care metrics in Senegal, Kenya and Tanzania, over the study period. Breaks were created at the 33rd and 66th percentiles for the two metrics.

Appendix C

APPENDIX: CHAPTER 4

C.1 Supplementary Data

Table C.1: List of covariates considered to model under-fiver mortality rates by subnational units

Covariate	Spatial resolution	Temporal resolution	Source
Malaria incidence	Admin-level	Annual	MAP [118]
LRI incidence	Admin-level	Annual	IHME [67]
Diarrhea prevalence	Admin-level	Annual	IHME [66]
MCV1 coverage[a]	Admin-level	Annual	IHME [105]
BCG1 coverage[a]	Admin-level	Annual	IHME
PCV3 coverage[b]	Admin-level	Annual	IHME [104]
Polio3 coverage[b]	Admin-level	Annual	IHME
Hib3 coverage[b]	Admin-level	Annual	IHME
Access to WASH	Admin-level	Annual	IHME [106]
Total population under five years old	Pixel-level	Annual	WorldPop

[a] % of children under one year of age who have received at least one dose of measles-containing vaccine (MCV1), or one dose of bacille Calmette-Guérin (BCG1) vaccine in a given year

[b] % of children ages 12–23 months who received three doses three doses of pneumococcal conjugate vaccine (PCV3), polio vaccine (Polio3), or Haemophilus influenzae type B vaccine (Hib3), in a given year.

C.2 Supplementary Methods

C.2.1 Model selection

Direct estimates and design-based variance estimates were computed using the ***survey*** package in *R*. We fit the Bayesian hierarchical models using the Integrated Nested Laplace Approximation [32] and the R-INLA package [33]. We obtained a subset of all included covariates by checking for multi-collinearity using the variance inflation factor (VIF) with a threshold of 5. For each indicator and country, we compared the 6 models presented in table C.2 consisting of different combinations of the covariates selected with the VIF procedure, and spatio-temporal random effects, using three selection procedures (the deviance information criteria [DIC], the Watanabe-Akaike information criteria [WAIC], and the conditional predictive ordinate [CPO]). In the absence of consensus on a single criterion, when different criteria pointed to different models, we used a majority rule.

Table C.2: Models considered to estimate the association between quality-adjusted coverage metric and under-five mortality rates, by subnational units and over time

Model	Formula
1	$\gamma_t + \alpha_t + e_i + \delta_{it}$
2	$\gamma_t + \alpha_t + e_i + S_i + \delta_{it}$
3	$X_{it}\boldsymbol{\beta} + \gamma_t + \alpha_t + e_i + \delta_{it}$
4	$X_{it}\boldsymbol{\beta} + \gamma_t + \alpha_t + e_i + S_i + \delta_{it}$

C.3 Supplementary Results

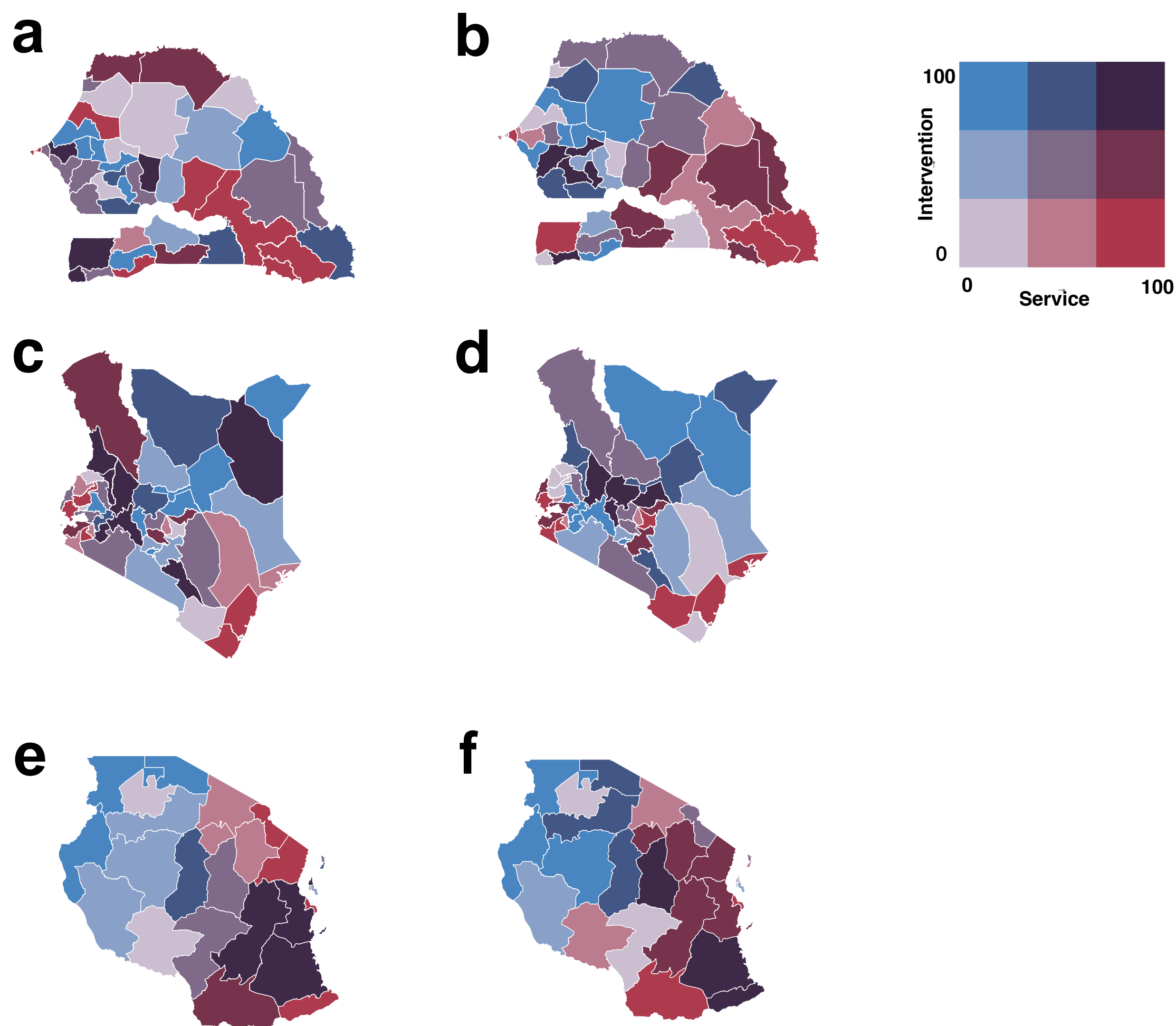


Figure C.1: Service and receipt of intervention coverage quantiles in 2010 (a) and 2020 (b) at the second administrative level in Senegal. Service and receipt of intervention coverage quantiles in 2010 (c) and 2020 (d) at the first administrative level in Kenya. Service and receipt of intervention coverage quantiles in 2010 (e) and 2020 (f) at the first administrative level in Tanzania.

C.3.1 Random effects summaries and visualizations

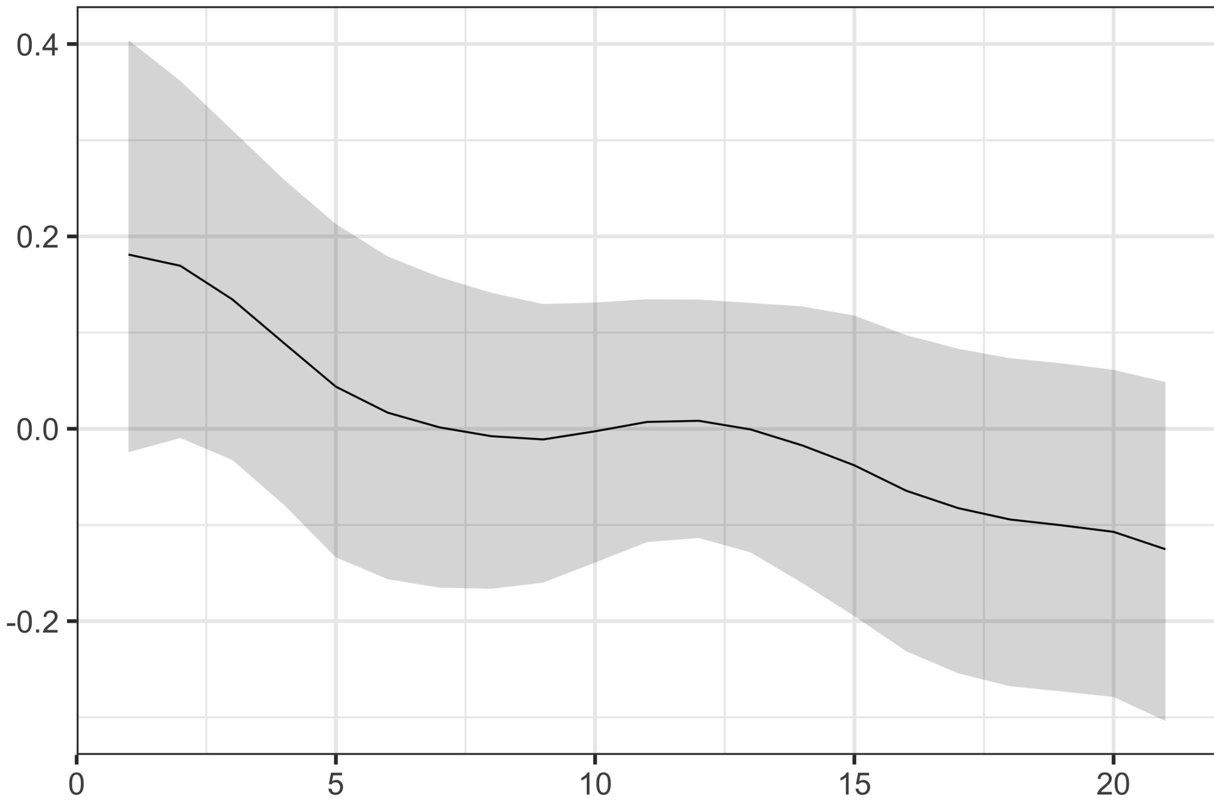


Figure C.2: Temporal random effects modelled as a first order random walk

C.3.2 Posterior distributions ecological regression model

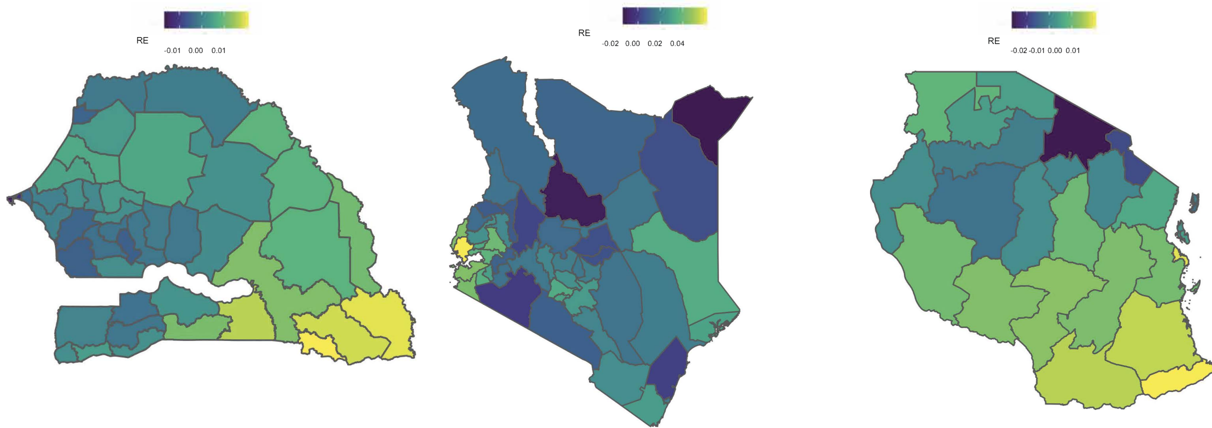


Figure C.3: Spatially correlated random effects

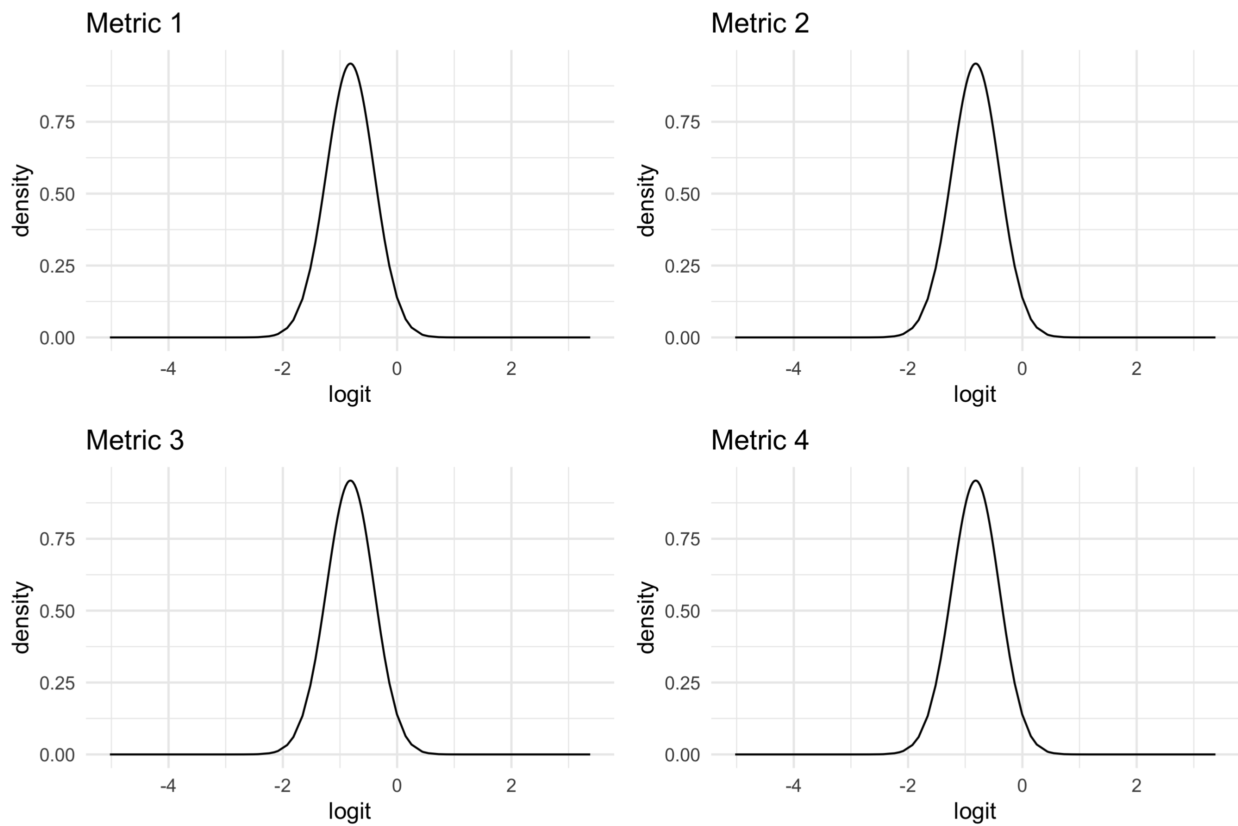


Figure C.4: Posterior densities of the fixed effects for quality-adjusted coverage under the four models (on the transformed-logit scale)

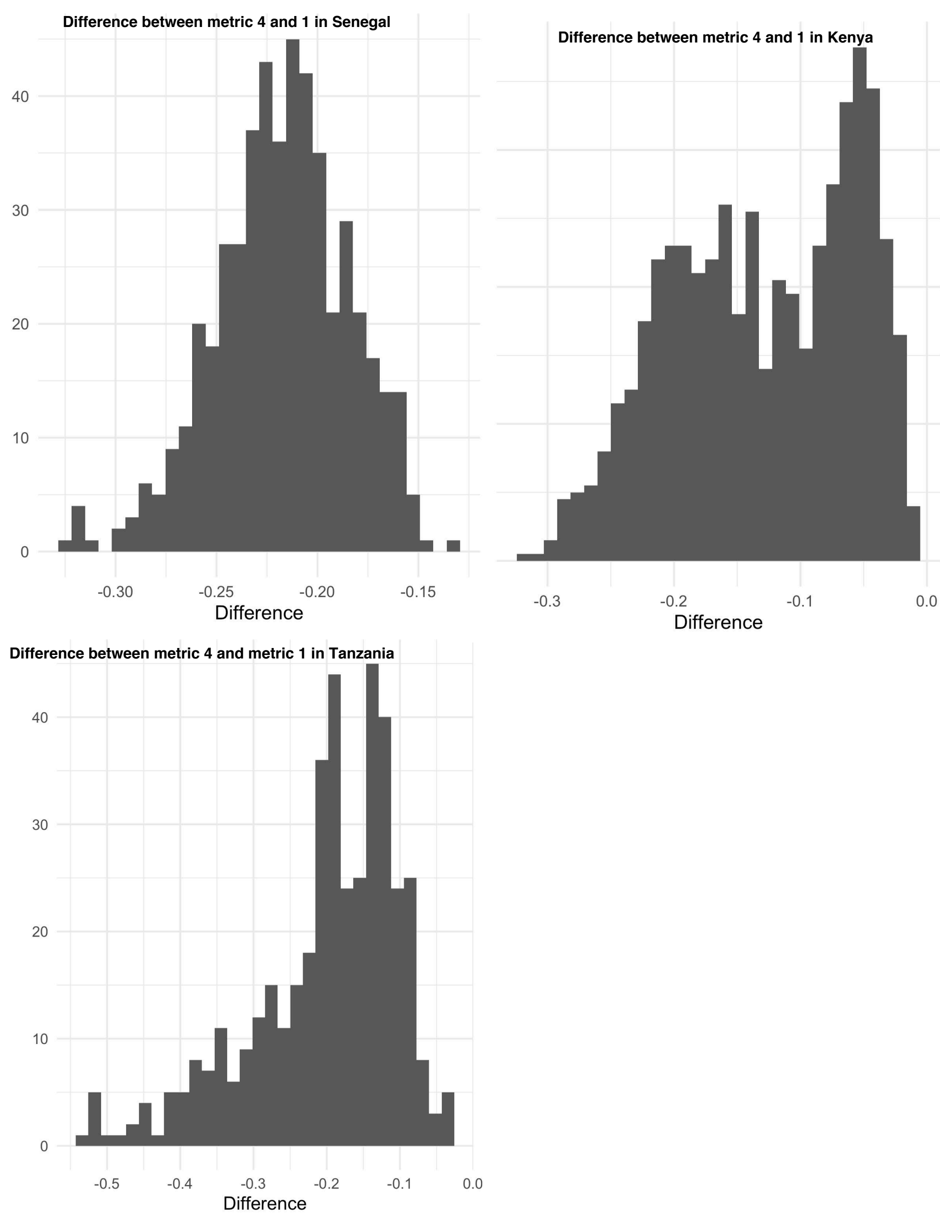


Figure C.5: Comparing posterior mean estimates of quality-adjusted coverage between metrics 4 and 1, in Senegal, Kenya, and Tanzania