NORTHWESTERN UNIVERSITY

Spatial and Temporal Methods to Analyze the Malaria Burden using Routine Health Facility Case Data in Burkina Faso

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Statistics

By

Sebastian Rodriguez

EVANSTON, ILLINOIS

August 2022

ABSTRACT

Spatial and Temporal Methods to Analyze the Malaria Burden using Routine Health Facility Case Data in Burkina Faso

Sebastian Rodriguez

Seasonal malaria chemoprevention (SMC) was first recommended by the World Health Organization (WHO) in 2012 to prevent uncomplicated malaria in children and began implementation in Burkina Faso in 2014 under programmatic campaigns. Systematic assessment of the impact of national SMC campaigns requires data with weekly or monthly temporal resolution over several years and broad spatial coverage. Despite the intensive deployment of malaria control interventions, such as SMC, health facilities in Burkina Faso reported increased malaria incidence between 2015 and 2018 among children under the age of 5. This analysis aims to adjust malaria routine case data in Burkina Faso for factors affecting the reported disease burden and describe the changes in the trend of malaria transmission between 2015 and 2018. The goal is to refine the subnational analysis of trends in the malaria burden in Burkina Faso and improve the assessment of SMC impact through quasi-experimental modeling.

Monthly data on confirmed malaria cases from 2015 to 2018 was obtained from Burkina Faso's Health Management Information System (HMIS). The data was aggregated to the health district level and was adjusted for changes in the health facility reporting rate and changes in treatment-seeking behavior. Seasonal trend decomposition was applied to crude malaria incidence, adjusted malaria incidence, all-cause outpatient visits, non-malarial outpatient visits, and the malaria proportion of outpatient visits to determine the trend in each time series for children under 5 years old and individuals over 5 years

old. The Sen's slope coefficient was used to quantify the trend components and map them for each health district in the country. Using adjusted malaria outcomes, we quantify the effectiveness of SMC using district-level data against symptomatic malaria with generalized linear mixed-effects models in a difference-in-differences framework. We quantify the effectiveness of SMC in reducing malaria incidence, the malaria proportion of outpatient visits, and the incidence of malaria hospitalizations for each group of health districts treated in the same year.

Given the impact of changes in treatment-seeking behavior spurred on by a national change in healthcare policy that removed healthcare fees for children under 5 years old and pregnant women, we modeled the treatment-seeking rate at a 1×1 km spatial resolution as a latent process. This latent process was modeled using data from the Malaria Indicator Survey from 2014 and 2017/18 with an integrated nested Laplacian approximation (INLA) using spatial random effects with a Matérn correlation structure. For analysis focused on the malaria burden at the health district level, fine-scale estimates were aggregated for each district using a population-weighted average.

For children under 5 years old, trends increased for crude malaria incidence (43/70 health districts). After adjusting for reporting rates and the changes in modeled treatment-seeking rate, adjusted malaria incidence trends increased in 26 districts, decreased in 4 districts, and remained unchanged in 40 districts. The proportion of outpatient visits that were confirmed malaria cases had a decreasing trend in 64/70 districts. For individuals over 5 years old, trends in malaria incidence and outpatient visits increased more modestly or remained unchanged in most districts.

SMC reduced the adjusted incidence in children under 5 years old by 27.1% (95% CI: [19.8%, 33.7%]), 43.4% (95% CI: [36.2%, 49.9%]), and 29.7% (95% CI: [20.2%, 38.1%]) in the $2016-2018$ SMC groups, respectively. SMC showed a weak protective effect on reducing the malaria proportion of outpatient visits in children under 5 years old: 7.5% (95% CI: [2.3%, 12.5%]) reduction for the 2016 SMC rollout group, 13.7% (95% CI: [7.4%, 19.5%]) reduction for the 2017 group, and no significant effect for the 2018 group. SMC was also associated with a reduction in malaria hospitalizations among children under 5 years old by 39.8% (95% CI: [31.4%, 47.2%]) in the 2016 SMC rollout group and 35.2% (95% CI: [20.5%, 47.2%]) in the 2017 group. SMC was not associated with a significant reduction in any of the three indicators

for individuals over 5 years old, except in 2017 where a reduction of 12.2% (95% CI: [1.6%, 21.6%]) was observed for adjusted malaria incidence and a reduction of 15.2% (95% CI: [1.2%, 27.2%]) was observed for malaria hospitalizations.

The gaps in reporting quality among healthcare facilities limit our ability to leverage the full HMIS data for the evaluation of the malaria burden and the estimation of the effectiveness of SMC. Using the HMIS data at the healthcare facility level, we address the gaps in reporting by expanding the analysis of the malaria burden among children under 5 years old and the effectiveness of SMC using a spatio-temporal model in an INLA framework. We model the malaria proportion of outpatient visits among children in a generalized linear mixed-effects model and use this to capture the effectiveness of SMC in a difference-in-differences framework. Time series are modeled from health facilities with known GPS locations and the model includes a temporal random effect with a seasonal correlation structure and a spatio-temporal random effect with a Matérn correlation structure. The modeled latent process for the treatment-seeking rate is included in the model to help adjust for changes in treatment-seeking behavior to each healthcare facility. We introduce a novel approach to difference-in-differences modeling frameworks, that helps capture the impact of SMC and changes in treatment seeking on the malaria burden. We map the malaria proportion of outpatient visits at a 1×1 km resolution for the 4 months of the high-transmission season for malaria each year in the study period. We derive the effectiveness of SMC using healthcare facility-level data, and compare the results to the model using district-aggregated data.

Using healthcare facility-level data in a spatio-temporal model, SMC reduced the malaria proportion of outpatient visits among children under 5 years old by 10.9% (95% CI: [8.3%, 13.5%]) in the 2016 SMC group, 16.2% (95% CI: [12.9%, 19.3%]) in the 2017 SMC group, and 11.0% (95% CI: [5.8%, 15.9%]) in the 2018 SMC group. The results of SMC effectiveness from the 2016 and 2017 groups are consistent with those found in the health district level model. This model was able to capture a significant reduction associated with the introduction of SMC in the 2018 SMC group, whereas the district-level model found no significant reduction.

SMC showed a mild to moderate protective effect against symptomatic malaria among children under 5 years old. Changes in treatment-seeking behavior and health facility reporting contributed to

the observed increase in malaria incidence in children under 5 years old between 2015 and 2018, but adjusting for these factors did not completely remove the increasing trend across districts. The choice of which malaria outcome to analyze shows strong differences in the protective effectiveness of SMC, as the 2018 rollout group saw a significant reduction in malaria among children under 5 associated with SMC but no such significant reduction in the malaria proportion of outpatient visits. The choice of using the malaria proportion as the outcome variable helps us control for monthly changes in treatment seeking, but does not directly reflect changes in malaria incidence. Using adjusted malaria incidence as the response variable, we cannot reliably control for intra-regional or temporal changes in treatment-seeking behavior. Using spatio-temporal modeling frameworks helps address some of the limitations in the routine case data caused by gaps in reporting quality. These frameworks helped in capturing the effectiveness of SMC on the 2018 SMC group, which was not observed to be significant using a district-aggregated model. Adjusting for sources of bias is a necessary step before analyzing routine data for SMC effectiveness. Our novel approach to differences-in-differences frameworks which include latent processes may be extended to other disciplines which utilize differences-in-differences modeling frameworks while considering latent processes.

Acknowledgements

This thesis required the support and collaboration from a diverse group of individuals scattered across the world. I would have liked to thank them all personally.

I would first and foremost like to thank my advisor and mentor, Dr. Noelle Samia, who guided me through challenging projects and situations with wisdom and kindness. There are few advisors who are as personable, supportive, and knowledgeable as her, and I feel truly fortunate for the professional relationship we developed as well as our friendship. She was always in my corner, looking out for my best interests and advocating for me at any point possible. She introduced me to members of the Institute for Global Health at Northwestern and allowed me to focus on applied research which I found meaningful and gratifying. I am eternally grateful for her patience, humor, and companionship throughout my years at Northwestern. Thank you for believing in me when it was tough to believe in myself.

Likewise, I would like to thank my co-advisor and mentor, Dr. Jaline Gerardin, who took me under her wing in her newly formed lab and gave me the opportunity to pursue meaningful applied research in the global health sector. She had an incredible amount of patience as I became acquainted with malaria and the literature surrounding it. She helped me develop professional relationships with collaborators around the world. I thank her for making me a part of her research group, which I had the privilege to see grow and take shape into a wonderful, international research community.

Thank you to the Northwestern Malaria Modeling lab, whose members have each been friends, mentors, and collaborators. There are too many members to list individually, but I would like to give a special thanks to Ifeoma Ozodiegwu who was a huge source of inspiration and a mentor throughout my time in the lab. I would also like to thank Kalamdeen Okuneye, who was the only member of the lab when I joined and helped me find my footing in the area of global health research when I was still in my

early Ph.D. years. A special thanks to the rest of the team, you have all been a wonderful group to work with, and I will dearly miss our weekly interactions.

I want to thank the Northwestern Department of Statistics, whose faculty and students have been a great source of learning, guidance, and companionship. I had the privilege of being part of a wonderful cohort with Abby Smith, Zhipeng Hou, Yajun Liu, and Frank Fineis. Thanks for being such great people, willing to help each other get through the first and most difficult times of our Ph.D. I feel like the COVID-19 pandemic cut our time together, and I would have loved to spend more time with you all. Thanks to all of the other students in the department with whom I shared a friendship. A special thanks to Mena Whalen for being a great friend and mentor. Thanks to Rrita Zejnullahi, Thomas Ippolito, and Yiben Yang for being around to share a laugh. Thank you to Kisa Kowal for being an infallible pillar in the department. Thank you to all the professors who imparted vast knowledge through classes and conversations. A special thanks to my committee member and teacher Hongmei Jiang, who kindly joined our group and supported my work.

We would like to give a special thanks to our partners at the Burkina Faso Programme National de Lutte contre le Paludism for sharing their data and support. Specifically, we thank Jean Pascal Sandwidi, Jean Baptist Ouedrougou, and Gauthier Tougir for maintaining and cleaning the HMIS data. Additionally, we thank our partners at the WHO Global Malaria Programme for their collaboration and support with our project. Special thanks to our collaborators and co-authors Beatriz Galatas and Abdisalan Noor. The findings and conclusions in this thesis are those of the author and do not necessarily represent the official position of the Burkina Faso PNLP or the WHO Global Malaria Programme.

I would also like to thank Dr. Melissa Simon and Dr. Joseph Fienglass for their support through the T37 Northwestern University Minority Health and Health Disparities Research Training Program. Research reported in this publication was supported by the National Institute On Minority Health And Health Disparities of the National Institutes of Health under Award Number T37MD014248. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Thank you to my undergraduate advisor, Dr. Harish Bhat, for pushing me to pursue a Ph.D. and opening the door for me to do research from an early stage in my education.

Lastly, a special and dear thanks to all the friends I made throughout my time here. Linda Pei, my Gemini twin, the light that guides me home, my sleep paralysis demon, thanks for being my roommate for three years. We made it through the COVID-19 pandemic without killing each other. You were a pillar of emotional stability for me, and I will miss you as we transition to the next chapter in our lives. Ranya Virk, thanks for being a fantastic friend and taking me on walks. I have huge admiration for your achievements and am so happy about the amazing opportunities coming up in your life. Reem Rashid, thanks for always being willing to make a trip to get fried chicken, the food of gods, with me. You were one of the first very close friends I made while at Northwestern and I am so happy that we will be able to maintain this wonderful friendship in the coming years of us living in Chicago. A special thanks to Abby Smith. You were truly my first close friend at Northwestern and opened the door for me to meet many other wonderful people. You have been a close confidant and a great person to lean on many times through the rough parts of our Ph.D. Thanks to Neff and Leo for being true homies, through bad times and good times, and for being around to hang out online.

A super special thanks to my beautiful and wonderful girlfriend, Yena Lee. You are a genius scholar and a beacon of joy in my life. Thanks for sticking with me through uncertain times and coming up with yummy ideas for food to cook together. Thanks to my siblings Nono y Luisi, for being huge motivators for me to achieve great things. Thanks to my parents and grandparents for their love and support. I wish I was physically closer to you all but am thankful for the amazing relationship that keeps us close emotionally. I love you all.

List of abbreviations

ACT: artemisinin-based combination therapy

CHW: community health worker

CI: confidence interval

CPO: conditional predictive ordinate

DHIS2: district health information software version 2

DHS: demographic & health survey

DIC: deviance information criterion

DiD: difference-in-differences

GF: Gaussian field

GLMM: generalized linear mixed-effects model

GTS: global technical strategy

HBHI: high burden high impact

HF: healthcare facility

HMIS: health management information system

INLA: integrated nested Laplacian approach

IRR: incidence rate ratio

IRS: indoor residual spray

ITN: insecticide treated bednets

LLINs: long-lasting insecticide treated bednets

MAE: mean absolute error

MAP: malaria atlas project

MDA: mass drug administration

MIS: malaria indicator survey

NMCP: national malaria control program

NMSP: national malaria strategic plan

ov5: individuals over 5 years old

PE: protective effectiveness

PNLP: Programme National de Lutte contre le Paludisme (Burkina Faso National Malaria Control Program)

RDT: rapid diagnostic test

SAE: small-area estimation

SLCPO: negative sum of the log CPO

SMC: seasonal malaria chemoprevention

STL: seasonal trend decomposition based on local (LOESS) regression smoothening

SPDE: stochastic partial differential equation

TS: treatment seeking

u5: children under 5 years old

WAIC: Watanabe–Akaike information criterion

WHO: world health organization

WMR: world malaria report

Dedication

I dedicate this thesis to my family, especially my parents and grandparents who have sacrificed great things for me to achieve what I have achieved.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

This project seeks to understand recent trends in the reported burden of malaria in the digital Health Management Information System (HMIS) of Burkina Faso and to model the effectiveness of seasonal malaria chemoprevention (SMC) using routine case data. Over the last decades, the District Health Information Software version 2 (DHIS2) data collection system, sometimes implemented as the HMIS, has been adopted in several sub-Saharan African countries with the purpose of bolstering disease surveillance systems across health facilities (World Health Organization, 2021). The effective use of surveillance data like the HMIS with high spatio-temporal resolution is seen as a vital intervention to combat the malaria burden and achieve eventual eradication worldwide (World Health Organization, 2015).

This chapter focuses on introducing the reader to the epidemiology of malaria and its impact on global health. The first section discusses the basics of the malaria lifecycle and the diagnosis, treatment, and interventions for the disease. In the third section, the high burden high impact initiative to combat malaria in countries with the highest burden is introduced. The fourth section in this chapter address different forms of surveillance data, limitations in their use, and gaps in understanding their behavior, as well as the global strategy of malaria eradication. The final section provides an overview of the setting in Burkina Faso and how the chapters in this project address the use of HMIS data to develop and evaluate malaria control strategies.

1.1. Background on malaria

Malaria is a mosquito-borne illness caused by the presence of parasites in the blood. Malaria is usually associated with fevers, chills, and other flu-like symptoms. Diagnosis and treatment are usually inexpensive and effective if sought promptly after symptom onset. However, if left untreated, severe

complications may develop in the form of severe anemia, acute respiratory distress syndrome, and cerebral malaria–categorized by impaired consciousness, seizures, or other neurologic anomalies–and could lead to death ("Fact Sheet about Malaria," 2022).

Malaria is a devastating disease, with an estimated 241 million cases and 627,000 deaths worldwide in 2020. A majority of deaths occur in children in the African continent, as children can have weaker immune systems and are more likely to develop severe morbidity from malaria. In 2020, children under 5 years old accounted for 80% of all malaria deaths in Africa ("Fact Sheet about Malaria," 2022).

The malaria lifecycle is typified by the cyclical infection of the female Anopheles mosquitoes, the vector population, and the humans bitten. A female mosquito will infect the human with the malaria parasites. These parasites grow and multiply, initially in the liver cells until affecting the red blood cells. An uninfected mosquito can then be infected by malaria parasites upon biting an infected human, thereby completing the cycle ("CDC - Malaria - about Malaria - Biology," 2020).

Diagnosis and detection of malaria infections in an individual are usually done with rapid diagnostic tests (RDTs) or microscopy-based techniques ("Fact Sheet about Malaria," 2022). While microscopy provides the gold standard of detection, RDTs are much cheaper and easy to execute by community health workers (CHWs) and other staff not trained to perform microscopy. RDTs can be conducted as part of household survey studies to assess malaria incidence in communities and for prompt detection of malaria in remote communities with limited access to care. An important limitation of RDTs is their propensity to detect antigens associated with malaria infections for a highly variable amount of time after an individual is treated with anti-malarials (Dalrymple et al., 2018).

Several interventions have been used globally to combat the burden of malaria. Simple control techniques such as insecticide treated bednets (ITNs) are highly effective ("Fact Sheet about Malaria," 2022) in reducing malaria infections during the nighttime when they are in use and have seen increased coverage across the African continent (World Health Organization, 2021). These bednets have been a part of mass intervention campaigns with a broad aim of national coverage. Similar campaigns have involved mass drug administration (MDA), in which drugs including chloroquine, primaquine, pyrimethamine, and sulfadoxine-pyrimethamine are distributed into communities to accelerate malaria eradication (Newby et

al., 2015; Poirot et al., 2013), although with mixed success due to the possibility of increased drug resistance. Interventions targeting areas of specific interest have been explored, with indoor residual spraying (IRS) campaigns applying insecticide coating to the interior of houses in certain communities (The PMI VectorLink Project, Abt Associates Inc., 2020; World Health Organization, 2021).

More recently, seasonal malaria chemoprevention (SMC) has been used in African countries, aiming at reducing the malaria burden in children. SMC is prophylactic therapy first recommended by the World Health Organization (WHO) in 2012 and given to children under 5 years old in areas categorized by high endemicity and high seasonality (World Health Organization and others, 2012). SMC is given during the peak malaria transmission period and the foremost recommended treatment is a single dose of sulfadoxine/pyrimethamine combined with a 3-day course of amodiaquine, once a month for up to 4 months. Meta-analysis from clinical trials has shown between $75-88\%$ reduction in clinical malaria with SMC (Cairns et al., 2021; Meremikwu et al., 2012; Wilson & Taskforce, 2011). Thus, the goal is universal coverage of children in the desired age-group to achieve a prophylactic effect during the ages at which severe malaria can be most dangerous.

Additionally, the RTS,S vaccine to prevent malaria has been adopted in 2019 in 3 sub-Saharan African countries: Ghana, Kenya and Malawi. In a study conducted over the course of 5 years, the RTS,S vaccine saw a 39% reduction in uncomplicated malaria cases and a 29% reduction in severe cases among children 5-17 months old who received 4 doses of the vaccine ("Malaria," 2020). These relatively low results for the protective efficacy of the RTS,S vaccine, along with high costs of development and distribution, has limited programmatic implementation of this intervention.

In the case of infection, the predominant way to treat malaria caused by the P.falciparum parasite is with artemisinin-based combination therapy (ACT). ACT is a fast acting and highly effective treatment that eliminates parasites from the blood ("Fact Sheet about Malaria," 2022). Due to the malaria lifecycle, malaria treatment also reduces transmission of infection by reducing the infectious reservoir for the vector population. Combination therapies have emerged from the concern of the rapid development of resistance to artemisinin-based monotherapy. Thus, the adoption of ACTs as the frontline treatment for P.falciparum has contributed towards malaria control and eradication. Outside of regions where

P.falciparum is the dominant parasite, chloroquine is the primary treatment for P.vivax and P.ovale, the other prevalent strains of the malaria parasite ("Fact Sheet about Malaria," 2022).

1.2. High burden high impact countries

According to the WHO, 70% of the malaria burden worldwide is concentrated in 11 countries. These are India and 10 countries from sub-Saharan Africa–Burkina Faso, Cameroon, Democratic Republic of the Congo, Ghana, Mali, Mozambique, Niger, Nigeria, Uganda and United Republic of Tanzania. Among these high-burden counties, there was an estimated 163 million malaria cases and 444,600 deaths (World Health Organization, 2021) in 2020. The Global Technical Strategy for Malaria 2016-2030 (GTS) set a goal of reducing malaria cases and deaths by at least 40% by 2020, at least 75% by 2025, and at least 90% by 2030, worldwide (World Health Organization, 2015). Between 2019 and 2020, the 10 highest burden African countries reported increases in malaria cases and deaths, signaling that the goals set by the GTS are unlikely to be met unless major policy changes are put in place (World Health Organization, 2021). This issue is further compounded by insufficient funding for malaria control and eradication efforts.

In 2018, the WHO and RBM Partnership to End Malaria launched the high burden, high impact (HBHI) initiative to support malaria reduction efforts in these countries. The HBHI initiative is a targeted malaria response to get countries back on track to achieve the GTS milestones (World Health Organization, 2021; World Health Organization and others, 2018a). A central goal of campaigns promoted by the HBHI initiative is that intervention planning and malaria control strategies is done sub-nationally to most efficiently deliver aid and most effectively allocate the limited resources (Galatas et al., 2021). Therefore moving away from broad campaigns with national distribution, such as MDA and mass ITN campaigns, towards sub-nationally tailored interventions informed by high resolution local data.

One of the core elements of the HBHI strategic plan is to update the previous malaria program reviews into a novel National Malaria Strategic Plan (NMSP) using readily available, structured data (Galatas et al., 2021). This would allow analysis to be tailored to the subnational regions of interest, capturing heterogeneity in demographics, environment, intervention history, and malaria transmission

rates. Surveillance data from country health systems provides one opportunity to leverage routine data to achieve the subnational tailoring of strategic plans.

1.3. Active and passive malaria surveillance data

National estimates of the malaria burden have typically relied on survey estimates from sources like the Demographic & Health Survey (DHS), the Malaria Indicator Survey (MIS), and other cross-sectional surveys. The DHS and MIS are household surveys conducted as part of active surveillance programs and provide national and regional estimates across a wide variety of indicators. National surveys like these can capture wide ranges of disease metrics throughout the country, enabling the collection of detailed data from households (Ozodiegwu et al., 2021). However, national surveys like these are often not powered to produce accurate estimates at a fine spatial resolution, limiting the ability for these surveys to be used for the subnational tailoring of interventions (Galatas et al., 2021). Furthermore, DHS/MIS surveys are only conducted every 3-4 years, meaning that these surveys are limited in tracking monthly or even yearly changes to the malaria burden. Active case surveillance methods for malaria that aim to estimate the malaria burden in communities with door-to-door surveys or cohort studies, are expensive to coordinate and lead to high quality data that is limited in spatio-temporal scope (Tiono et al., 2014).

Passive case surveillance for malaria involves the estimation of incidence from reported malaria cases and deaths occurring at health centers, or other community centers, observing and confirming malaria cases. The most ubiquitous system for digital passive case surveillance in sub-Saharan Africa is the District Health Information Software version 2 (DHIS2), commonly implemented as a health management information system (HMIS). Passive surveillance systems provide an essential dataset for the surveillance of the malaria burden, as these systems provide routine reports collected at a fine temporal resolution from health facilities across the nation or region of interest (Alegana et al., 2020; Tiono et al., 2014). Thus, datasets like the DHIS2 can be highly effective to tailor subnational strategies and combat the malaria burden. Limitations in passive surveillance systems include non-malarial factors, such as health policy changes and interruptions in reporting, affecting the monthly reported malaria counts. Gaps in reporting are common in areas of turmoil and conflict (Alegana et al., 2020; Tiono et al., 2014). For

example, HMIS data from 2019 in Burkina Faso is unavailable due to a workers' strike from the data collection staff (Rouamba et al., 2020). Crucially, these datasets only track reported malaria at health centers. Thus, malaria treatment sought from traditional healers or cases and deaths not observed in the health sector may be missed entirely. This would most affect rural communities without easy access to health systems, and thus could produce a substantial gap in certain demographic groups.

Pillar 3 of the GTS for malaria calls for the transformation of malaria surveillance into a core intervention, underscoring the importance of understanding the gaps in the surveillance systems in place and how they can complement each other to provide a more complete understanding of the malaria burden (World Health Organization, 2015). To best understand the malaria burden at a subnational level, datasets covering several years with high temporal resolution and large spatial coverage are needed. These datasets present opportunities for spatio-temporal modeling to capture the spatial heterogeneity in factors affecting the local malaria burden (Cameron et al., 2021; Danwang et al., 2021; Nguyen et al., 2020; Rouamba et al., 2020). Therefore, the transformation of HMIS data into a core source of estimates for the malaria burden and understanding the limitations of this dataset is key to the subnational tailoring of interventions to be on track for the goals of malaria reduction set out by the GTS.

1.4. Transforming HMIS data in Burkina Faso into a core dataset

In 2020, Burkina Faso was the country with the 6th largest malaria burden among the HBHI countries, with roughly 8.2 million cases estimated that year (World Health Organization, 2021). Burkina Faso began implementing the HMIS in 2013 to capture routine malaria case reporting at health facilities throughout the country (Rasmussen, 2018). This dataset tracked the monthly case reports from over 2,900 health facilities from the public and formal private health sector. The HMIS data provides broad coverage of the monthly malaria burden across the health facilities in the country. One possible limitation of this kind of passive surveillance data is that populations in rural areas of Burkina Faso have historically sought care for malaria at home (Müller et al., 2003), therefore data on the malaria burden from health facilities in rural locations may greatly differ from the malaria burden in the community.

As this surveillance system was adopted in the country, the Burkina Faso National Malaria Control Program (Programme National de Lutte contre le Paludisme, PNLP) scaled up several interventions to control the malaria burden. Long-lasting insecticide treated bednets (LLINs) were part of a mass distribution campaign, with widespread net distribution taking place every 3 years beginning in 2010. Healthcare became free for children under 5 years old and pregnant women in 2016, aiming to increase treatment seeking among vulnerable populations (Ridde & Yaméogo, 2018). Between 2017 and 2019, IRS was used in households in 3 health districts (The PMI VectorLink Project, Abt Associates Inc., 2020). Lastly, SMC campaigns began rollout of the drug starting with 7 pilot health districts in 2014 and expanding to cover the whole nation by 2019 (World Health Organization and others, 2012).

Despite a history of intense interventions in Burkina Faso, estimated malaria incidence has increased over the last years (Rouamba et al., 2020; World Health Organization, 2021). Routine case data on confirmed malaria cases from the HMIS shows increasing counts of malaria incidence (Rouamba et al., 2020), calling into question the effectiveness of recent interventions. However, estimated incidence using routine case data can be biased by recent interventions influencing the trend of reported malaria.

This project addresses the gap in understanding malaria incidence in the Burkina Faso HMIS data to enable statistical modeling to capture the effectiveness of SMC in Burkina Faso. In Chapter 2, we analyze the long-term trends in reported malaria incidence in the HMIS using seasonal trend decomposition and the Sen's slope coefficient. We find that reported malaria cases in Burkina Faso may be increasing in part due to increases in reporting rate from health facilities and improved treatment-seeking behavior from individuals, and we apply adjustments to control for the influence of these non-malarial factors. Chapter 3 presents an analysis of the effectiveness of SMC under programmatic implementation in Burkina Faso using data aggregated at the health district level, and builds on Chapter 2 by using adjusted data which has been controlled for sources of bias. We find that SMC has had a mild to moderate effect in reducing the malaria burden among children under 5 years old across health districts in the country. Lastly, Chapter 4 presents a fine-scale spatio-temporal statistical model built on data from geographically tagged healthcare facilities that captures the seasonal and spatial heterogeneities in the malaria burden. This model includes a latent process to estimate the changes in the treatment-seeking behavior at a

sub-district level, and approaches the differences-in-differences framework in a novel manner to estimate the effectiveness of SMC with the lingering effect of changes in treatment seeking. We produce smoothed maps of the malaria burden at a 1×1 km scale to visualize the impact of SMC. With this model, we find that our results for the impact of SMC using the most granular data provided by the HMIS are robust compared to the results found in Chapter 3.

CHAPTER 2

Quantifying trends in malaria incidence using routine case data in Burkina Faso in the presence of improved reporting quality and treatment-seeking behavior

2.1. Background

Burkina Faso is one of the countries with the highest malaria burden in the world, with an estimated 387 confirmed cases per 1000 population at risk in 2019 (World Health Organization, 2020). Children under 5 years old bear the greatest burden, accounting for 49% of all malaria cases and 74% of all deaths attributable to malaria in 2018 (Ministere de la Sante du Burkina Faso, 2019).

Over the last decade, Burkina Faso's National Malaria Control Program (Programme National de Lutte contre le Paludisme, PNLP) scaled up key malaria control interventions, many of which targeted children under 5 years old. Long-lasting insecticide-treated bednets (LLINs) were distributed in mass campaigns every three years beginning in 2010. Seasonal malaria chemoprevention (SMC) deployment began in 2014 in 7 pilot districts, and all 70 health districts were receiving SMC by 2019. Indoor residual spraying (IRS) was implemented in 3 districts between 2017 and 2019 (The PMI VectorLink Project, Abt Associates Inc., 2020). To expand access to healthcare for the most vulnerable populations, treatment at public health facilities became free of charge for children under 5 years old and pregnant women in April 2016 (Ridde & Yaméogo, 2018). In 2013, Burkina Faso began using District Health Information Systems 2 (DHIS2) as the digital Health Management Information System (HMIS) to capture routine malaria case reporting at health facilities (Rasmussen, 2018).

National malaria control programs are increasingly turning to routine case data as a valuable source of information for tailoring interventions at the subnational level. Despite the scaleup of interventions,

reported malaria incidence in Burkina Faso also increased from $2015-2018$ among children under the age of 5 (Rouamba et al., 2020; World Health Organization, 2020). While routine case data should in theory be able to capture long-term trends in malaria burden, routine data is also subject to changes in reporting practice, treatment-seeking, and other factors that may obscure temporal trends (Alegana et al., 2020; De Savigny & Binka, 2004). It is essential to adjust for underlying factors affecting the interpretability of routine case data when using this data source to evaluate trends in malaria transmission, stratify health districts for intervention allocation, and assess the impact of control interventions.

The High Burden to High Impact (HBHI) response was launched in Burkina Faso in May 2019 in partnership with the World Health Organization (WHO), the RBM Partnership to End Malaria, and other partners to support the PNLP in getting Burkina Faso back on track to achieve the 2030 Global Technical Strategy for Malaria targets (World Health Organization, 2020). During the review of their national malaria strategic plan, the PNLP applied an analytical process with the support of WHO and partners to tailor interventions to its 70 health districts and to design strategic and fully funded intervention mix plans appropriate to each district. Routine case data was used to stratify districts according to incidence in 2018, by adjusting the number of confirmed cases for testing and reporting rates at the district level to control for some of the known biases in this data (Galatas et al., 2021). The subnational tailoring of interventions exercise led to an updated strategic plan and a successful funding request to the Global Fund but raised several key questions around the use of routine data, including 1) how to measure changes in transmission as a result of the previous strategic planning cycle, and 2) how to evaluate the impact of specific interventions such as SMC introduced in Burkina Faso to reduce incidence.

In this study, we analyze recent trends in reported confirmed malaria cases, all-cause outpatient visits, and non-malarial outpatient visits in Burkina Faso at the health district level to understand changes in malaria transmission between 2015 and 2018 using routinely collected data and a refinement of the adjusting approach previously used during strategic planning.

2.2. Methods

2.2.1. Data sources and imputation

Monthly routine case data from public and formal private health facilities were extracted and validated by the PNLP from the Burkina Faso HMIS, covering 2,959 health facilities from January 2015 to December 2018, in the context of the subnational tailoring of interventions analysis that was part of the Burkina Faso HBHI response. Data fields were analyzed separately for children under 5 and individuals over 5 years old. Confirmed malaria cases were defined as those confirmed by microscopy ($< 0.3\%$) or rapid diagnostic test (RDT) ($> 99\%$). No correction was made for unconfirmed malaria cases, as presumed cases were not recorded prior to early 2016, and treated or suspected malaria cases were not consistently recorded. Thus, estimating presumed cases from excess untested or treated cases and the test positivity rate was not possible for the entirety of 2015 and over a third of the remaining data.

2.2.2. Analysis on the under-5 age group

Each row of the database corresponded to one health facility's monthly reporting. Among children under 5 years old, confirmed cases by RDTs, confirmed cases by microscopy, and all-cause outpatient visits were missing from 33.6%, 0.5%, and 20.2% of rows, respectively. Observations were imputed from health facilities missing no more than 6 observations in the time series and no more than 2 consecutive missing values. Missing values were imputed with a seasonally decomposed moving average method via the **imputeTS** R package (Moritz & Bartz-Beielstein, 2017). Imputation recovered 2,946 observations of confirmed RDT cases (6.2% of missing observations) from 1,314 unique health facilities, 738 observations of confirmed cases from microscopy (94.6% of missing observations) from 646 unique health facilities, and 898 observations of all-cause outpatient visits (3.1% of missing observations) from 500 unique health facilities.

Rows that met all the following inclusion criteria after imputation were retained: 1) the number of confirmed malaria cases was not missing; 2) the number of all-cause outpatient visits was not missing; 3) the number of all-cause outpatient visits was at least the number of confirmed malaria cases. This

process removed 45,628 out of 142,032 monthly records (32.1%) and 495 out of 2,959 (16.7%) health facilities entirely (Figure A.2). The cleaned health facility data were aggregated into 70 health districts according to their assignment in the HMIS database.

Annual population data for each health district was obtained from the PNLP. The fraction under 5 years old was estimated at 18% ("World Development Indicators," 2021) for all districts and years between 2015 and 2018.

2.2.3. Adjustments on crude incidence

Crude incidence was defined as the number of monthly confirmed malaria cases per 1000 children under 5 years old in the district. The monthly number of confirmed cases was adjusted at the health district level by reporting rate and treatment-seeking behavior based on a modified version of one of the approaches used in the World Malaria Report (WMR) by the WHO (World Health Organization, 2020). WHO's approach is based on number of reported confirmed cases and 1) accounts for unconfirmed presumed cases by adding the presumed cases multiplied by the test positivity rate, 2) accounts for unreported cases by dividing by the reporting rate, and 3) accounts for incomplete treatment-seeking by dividing by the treatment-seeking rate from the public and private health sectors. Since presumed cases or estimates for presumed cases were not complete, this analysis only considered adjusting for reporting rate and treatment-seeking rates (Equation A.1.4).

Not all health facilities were active during the entire period of $2015 - 18$. In the absence of a master list to provide dates of health facility activation or deactivation, we assumed that a health facility was inactive until it began reporting either confirmed malaria cases or all-cause outpatient visits, and we assumed a health facility became inactive if it did not report both confirmed malaria cases and all-cause outpatient visits for at least the last 6 months of the study period. By this methodology, 502 health facilities became active during the study period and 154 became inactive.

Monthly crude (unweighted) reporting rate was defined as the fraction of active health facilities in each health district that met the inclusion criteria each month. The monthly weighted reporting rate for each health district was determined as follows. For each calendar month (January through December),

health facility weights were calculated by dividing the health facility's average number of malaria cases reported for that month, across all years of data, by the district sum of the average number of malaria cases reported for that month, across all years of data. For dates in which a health facility is inactive, the average number of malaria cases reported for that month-and-year pair would be 0, and the weights for the health facilities in that district for that date would be calculated as usual. The district monthly weighted reporting rate was then calculated by summing the weights of the active health facilities that met the inclusion criteria each date (Equation A.1.3). This value captures the proportion of expected confirmed malaria cases that are reported at the district level each month by the active health facilities included in the HMIS database. The weighted reporting rate for all-cause outpatient visits among children under 5 was also calculated with this method. The adjusted incidence accounting for changes in reporting rate was calculated by dividing the crude number of confirmed cases by the weighted reporting rate (Equation A.1.5). After applying the reporting adjustments, we calculated the adjusted incidence per 1000 children under 5 years old in the districts by dividing by the under-5 population.

2.2.4. Small-area estimation (SAE) of district-level treatment-seeking rates with INLA

Treatment-seeking rates were derived from the 2014 and 2017-18 Malaria Indicator Surveys (MIS) and were defined as the fraction of children under 5 years old with fever in the last two weeks who sought medical treatment in the public or formal private health sector (ICF International, 2015, 2019). Bayesian spatial integrated nested Laplacian approximation (INLA) models were used to interpolate MIS cluster values and construct district-level estimates using a small-area estimation (SAE) method (Ingebrigtsen et al., 2014; Lindgren & Rue, 2015). Detailed SAE methods and results are described in Additional File A.2.

For each survey year, a generalized linear mixed-effects model (GLMM) was used with a Binomial distribution and a logit link function. The model included independent structured and unstructured spatial random effects, and climate predictors and travel-time to the nearest health facility as fixed

effects. An INLA model with a stochastic partial differential equations (SPDE) specification was used with the R-INLA package in R (Lindgren & Rue, 2015; Rue et al., 2009). The Matérn covariance function was used for the structured spatial random effect to introduce a measure of spatial autocorrelation (Rue et al., 2009). The unstructured, district-level random effect had a general Gaussian distribution and was included to account for inter-district heterogeneities. The climate predictors were average monthly precipitation of the previous month and average monthly surface-air temperatures of the previous month (Hersbach et al., 2020). The walking travel time to nearest health facility was calculated from PNLP's list of health facility locations and the friction surface in (Weiss et al., 2019). The selection of prior information for the model hyperparameters followed the standard specification within the R-INLA package (Krainski et al., 2018; Lindgren & Rue, 2015; Rue et al., 2009).

Treatment-seeking rates were modeled at a 1×1 km spatial resolution and aggregated to the district level using a population weighted average of the pixel-level posterior mean values, with population surface from WorldPop (Linard et al., 2012).

Monthly treatment-seeking rates were interpolated between MIS time points in three ways: 1) using the fitted values from a linear regression from the modeled 2014 estimate assigned to Aug 2014 to the modeled 2017-18 estimate assigned to Dec 2017; 2) using a step-function with the step in June 2016, with the modeled 2014 estimate assigned before the step and the modeled 2017-18 estimate assigned after the step; 3) using a step-function with the step in January 2017, with the modeled 2014 estimate assigned before the step and the modeled 2017-18 estimate assigned after the step. The step in June 2016 aligns most closely with when treatment of childhood illnesses became free of charge, while the January 2017 step was chosen since it is the midpoint of the time series. The adjusted incidence accounting for changes in treatment-seeking rate was calculated by dividing the number of cases adjusted for weighted reporting rate by each month's estimated treatment-seeking rate (Equation A.1.6), and then dividing by the under-5 population.

Monthly non-malarial outpatient visits among children under 5 years old was defined as the number of all-cause outpatient visits minus the number of confirmed malaria cases for each month and district, after each being adjusted for their respective weighted reporting rates.

The malaria proportion of outpatient visits among children under 5 years old was calculated by dividing the number of confirmed malaria cases by the number of all-cause outpatient visits, each adjusted for its respective weighted reporting rate, for each health district and month, and is bounded between 0 and 1.

2.2.5. Quantifying trends

To separate the overall temporal trend from the seasonal component of each indicator variable, seasonal trend decomposition based on local (LOESS) regression smoothening (STL) (Cleveland et al., 1990) was applied using the R package **stlplus** (Hafen, 2016; R Core Team and others, 2013). For each of Burkina Faso's 70 health districts, STL was applied to crude malaria incidence, malaria incidence adjusted for reporting rates, malaria incidence adjusted for reporting rates and each of 3 possible treatment-seeking increase profiles, all-cause outpatient visits per 1000 adjusted for reporting rate, non-malarial outpatient visits per 1000, and malaria proportion of outpatient visits. To ensure comparability between the indicator variables, each variable's time series was standardized by subtracting the mean value and dividing by the standard deviation. STL additively decomposes each time series Y_t into seasonal (S_t), trend (T_t), and remainder (R_t) components:

$$Y_t = S_t + T_t + R_t$$

The seasonal component describes the variation in the time series due to seasonal oscillations. The trend component describes the change in the direction of the data observed throughout the time series once the seasonal component has been removed. The values of the trend components do not have a specific interpretation, but their shape and direction in the time series inform about the long-term fluctuations in the monthly data. The remainder component collects the portion of Y_t that is left unexplained by S_t and T_t.

Sen's slope (Gilbert, 1987; Pohlert, 2016) was used to quantify the direction and intensity of the trends in the time series of each indicator variable and compare the district values across the different

indicators to assess changes after applying adjustments. Statistical significance of the trend was assessed by two-sided Mann-Kendall test (Gilbert, 1987; Pohlert, 2016).

2.2.6. Analysis on the over-5 age group

Routine data for individuals over the age of 5 was analyzed with the same methodology as for children under 5. Data for this age group was imputed via the same procedure and cleaned with the same inclusion criteria as described for children under 5. Trends were quantified for crude malaria incidence, malaria incidence adjusted for weighted reporting rate, all-cause outpatient visits, non-malarial outpatient visits and the malaria proportion of outpatient visits. Malaria incidence in this age group was not adjusted for changes in treatment seeking as there was no available data for changes in febrile treatment seeking rates for individuals over 5 years old.

2.3. Results

The number of active health facilities and the total number of facilities reporting data in Burkina Faso increased between January 2015 and December 2018 (Figure 2.1A). Reporting quality varied between regions (Figure 2.1B, 2.1C). The fraction of monthly health facility reports that met inclusion criteria increased between 2015 and 2018 for most regions, reaching over 90% in some regions, but remained persistently lowed for the Centre and Haut Bassins regions. The major urban areas of Ouagadougou and Bobo-Dioulasso are in Centre and Haut Bassins regions respectively, and the lower reporting rate may reflect a greater abundance of private health facilities in urban areas. Most exclusions were caused by missing data for confirmed malaria cases, but all-cause outpatient visits were often also not recorded when malaria cases were not recorded. Requiring all-cause outpatient visits in the inclusion criteria did not greatly affect the number of observations excluded per month.

The unweighted reporting rate shows some districts with below 50% health facility reporting (Figure 2.1D). After accounting for the relative caseload of each health facility in each district, reporting rate improved (Figure 2.1E), indicating that smaller health facilities with less patient traffic had the most difficulty with consistent reporting.

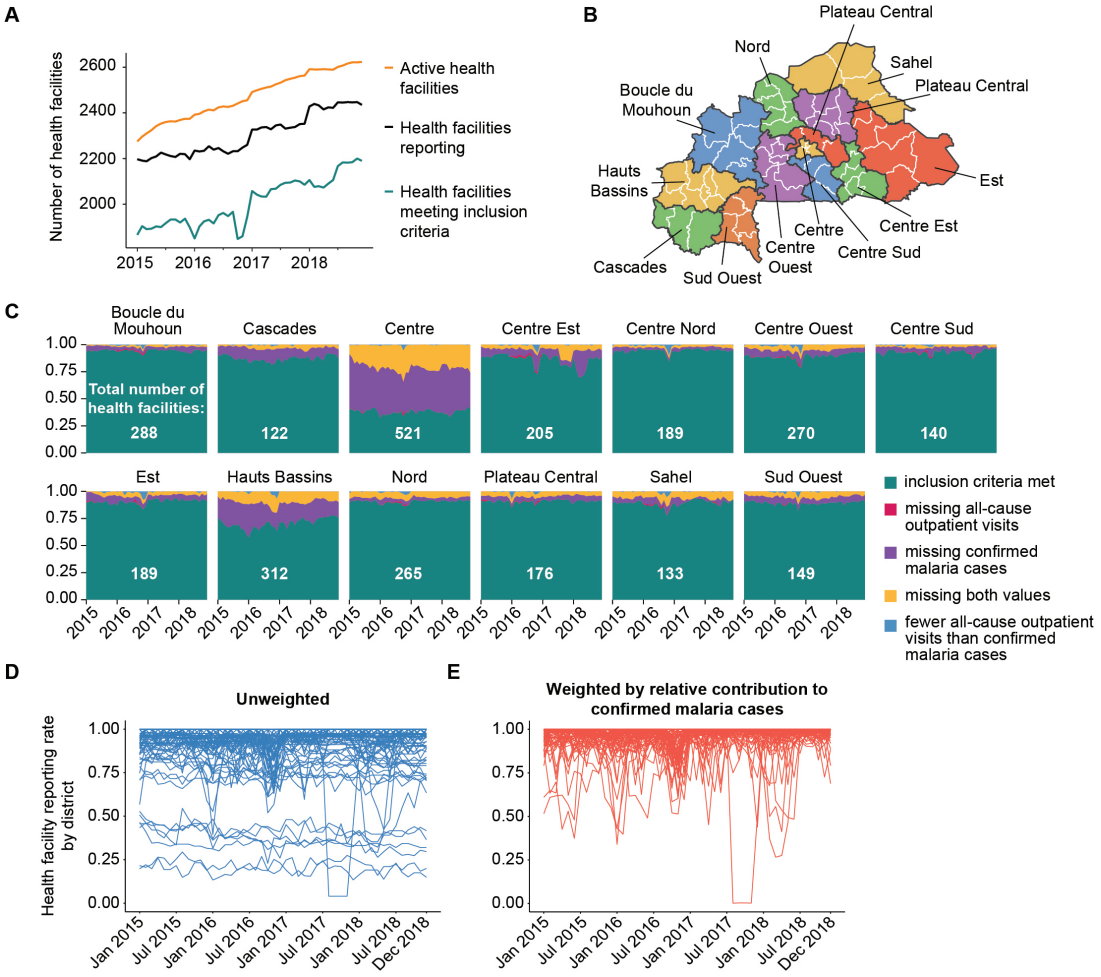


Figure 2.1. Reporting rate at health facilities in Burkina Faso, 2015-18, for data from children under 5 years old. A) Monthly number of active health facilities, number of health facilities reporting, and number of health facilities with reports meeting inclusion criteria. B) Administrative regions of Burkina Faso. Health districts are outlined in white. C) Regional monthly health facility reporting quality by fraction of reports meeting inclusion criteria or failing to meet each criterium. The total number of unique health facilities in each region is indicated in white. D) District monthly fraction of health facilities meeting the inclusion criteria (reporting rate). E) District monthly fraction of health facilities meeting the inclusion criteria, weighted by each health facility's expected contribution to the district's reported confirmed malaria cases.

Crude malaria incidence in children under 5 increased in Burkina Faso between 2015 and 2018 (Figure 2.2A, 2.2B, 2.2C). Most of the increase occurred in early 2016, around the time when children under 5 became eligible for free health care, and in late 2018 (Figure 2.2B). Among districts with a significant trend (44/70 districts), 43 had a positive Sen's slope coefficient, indicating an increasing trend (Figure

2.2C). Only Dori health district in the Sahel region had a significantly decreasing trend. After adjusting for the weighted reporting rate (Figure 2.2D), the trend in the adjusted incidence continued to exhibit a general upward trajectory in most districts (Figure 2.2E). The sign on the Sen's slope coefficient did not change for most of the districts but the value decreased (Figure 2.2F). In 2 health districts, the monotonically increasing trend in crude incidence became non-significant after adjusting for changes in reporting rate. Two health districts (Dori and Sebba in the Sahel region) had a decreasing trend in adjusted incidence. Results using the unweighted reporting rate show similar results as crude malaria incidence (Figure A.3).

Increases in treatment-seeking during the study period (Figure 2.3A) could be partially responsible for the increasing trend in malaria incidence. We considered three different functions (linear increase, June 2016 step increase, and January 2017 step increase) (Figure 2.3B) to capture the change in treatment-seeking rate at the regional level observed between the 2014 and 2017-18 MIS, since the specific timing of the change is not known. Example trends in malaria incidence in children under 5 after adjustment by each of the three possible changes in treatment-seeking are shown for Pama health district in the Est region (Figure 2.3C). Adjusting for treatment-seeking slightly flattened Pama's malaria incidence trend, and the largest change was observed when the increase in treatment-seeking was assumed to take place in a stepwise fashion in January 2017.

Across the 70 health districts, trends in malaria incidence in children under 5 after adjusting for treatment-seeking flattened (Figure 2.3). The increasing trend in early 2016 was still evident under the linear increase assumption, was most attenuated under the June 2016 step increase assumption, and remained high for many districts under the January 2017 step increase assumption (Figure 2.3D). After adjusting for treatment-seeking under the linear increase assumption, the coefficient for the Sen's slope indicated an increasing trend in 26/70 districts, decreasing trend in 5/70 districts, and a non-significant direction of the trend component according to the Mann-Kendall test in 39/70 districts (Figure 2.3E). Under the June 2016 step increase assumption, the Sen's slope coefficient showed an increasing direction of the trend component in 26/70 districts, a decreasing trend in 4/70 districts, and no significant direction of the trend component in 40/70 districts. Under the January 2017 step increase assumption, the Sen's

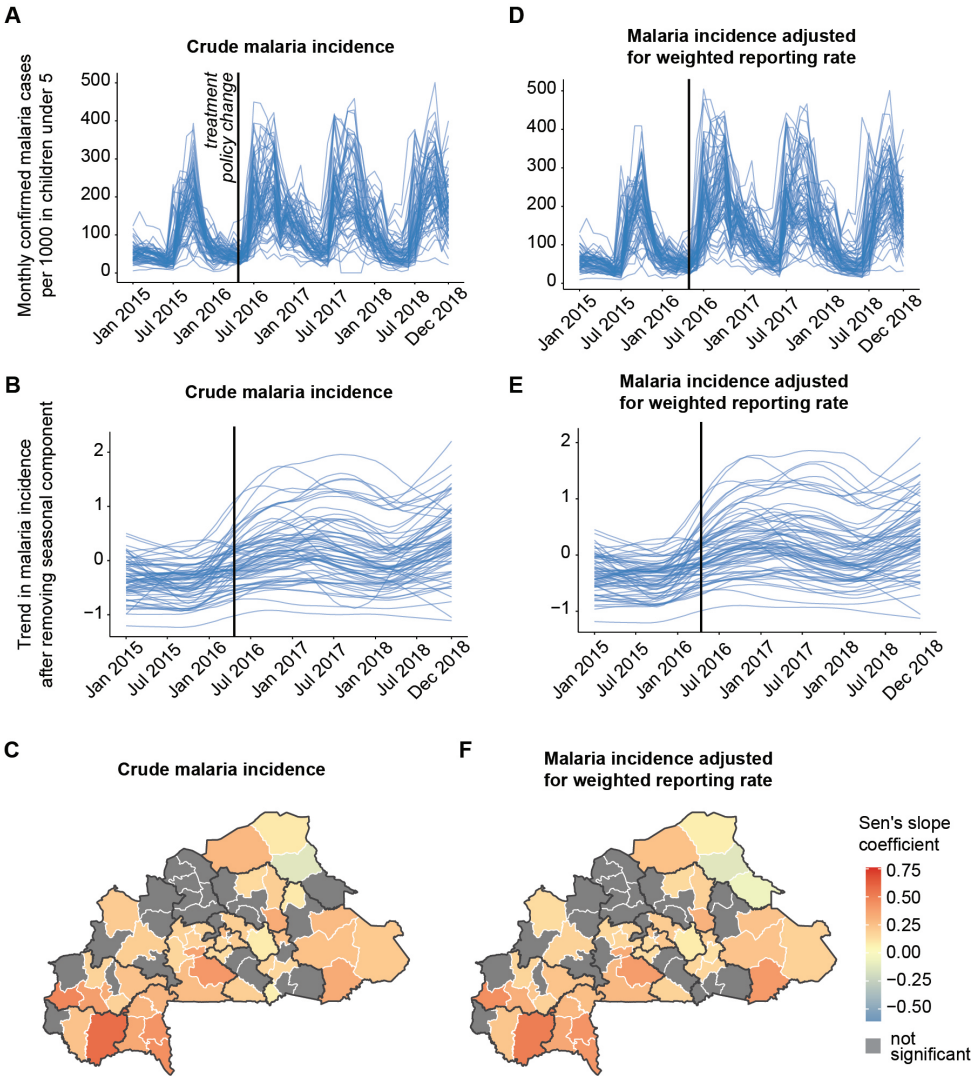


Figure 2.2. Quantifying trends in crude malaria incidence and malaria incidence after adjusting for weighted reporting rate in children under 5 years old. A) Monthly confirmed malaria cases per 1000 in children under 5 years (crude incidence). Each line represents one health district, and the black line indicates the 2016 policy change when treatment for childhood illnesses at public health facilities became free of charge. B) District-level trend components of crude incidence after seasonal trend decomposition. C) Sen's slope coefficient for the trend of crude malaria incidence. Positive numbers indicate increasing trend. Gray: no significant increasing or decreasing trend according to the Mann-Kendall test. D) Monthly confirmed malaria cases per 1000 children under 5 years after adjustment for weighted reporting rate. E) District-level trend components of malaria incidence adjusted for weighted reporting rate. F) Sen's slope coefficient for the trend of malaria incidence adjusted for weighted reporting rate.

slope coefficient was increasing for 19/70 districts, decreasing for 5/70 districts, and was non-significant for 46/70 districts. Districts in the Cascades, Sud Ouest, Hauts Bassins, and Sahel regions saw the largest changes in Sen's slope coefficients, reflecting their large changes in treatment-seeking (Figure 2.3A). In districts with significant direction in trend, Sen's slope coefficient tended to remain positive after adjusting for treatment-seeking in southern districts but become negative in northern districts.

All-cause outpatient visits in children under 5 increased for 63/70 districts between 2015 and 2018 (1/70 decreased, 6/70 not significant) (Figure 2.4A). The increase in all-cause visits was not only due to increase in confirmed malaria cases, as the increasing trend remained after subtracting confirmed malaria cases from all-cause visits (Figure 2.4B). Non-malarial outpatient visits in children under 5 increased in 66/70 districts over the study period (1/70 decreased, 3/70 not significant). Dori health district in the Sahel region was the only district with a negative trend in all-cause visits or non-malarial visits. The trend components of the all-cause and non-malarial visits timeseries showed the largest increase in early 2016, coincident with the change in healthcare policy for children under 5. Sen's slope coefficients of trends in both all-cause outpatient visits and non-malarial outpatient visits were higher than those for malaria incidence without adjusting for treatment-seeking (Figure 2.2F), suggesting that all-cause and non-malarial visits increased more between 2015 and 2018 than malaria incidence did. As a proportion of all-cause outpatient visits, confirmed malaria cases at healthcare facilities decreased over the study period for 64/70 districts (0/70 increased, 6/70 not significant) (Figure 2.4C), although the trend reversed in 2018 for many districts.

The increasing trends observed in 2016 of crude malaria incidence, malaria incidence adjusted for weighted reporting rates, all-cause outpatient visits, and non-malarial outpatient visits among children under 5 years old were not apparent for individuals over the age of 5 (Figure 2.5). Malaria incidence, all-cause outpatient visits and non-malarial outpatient visits all generally increased, with the largest increase observed among non-malarial outpatient visits. The malaria proportion of outpatient visits showed no significant trend for 42/70 districts (25/70 decreasing, 3/70 increasing) (Figure A.4).

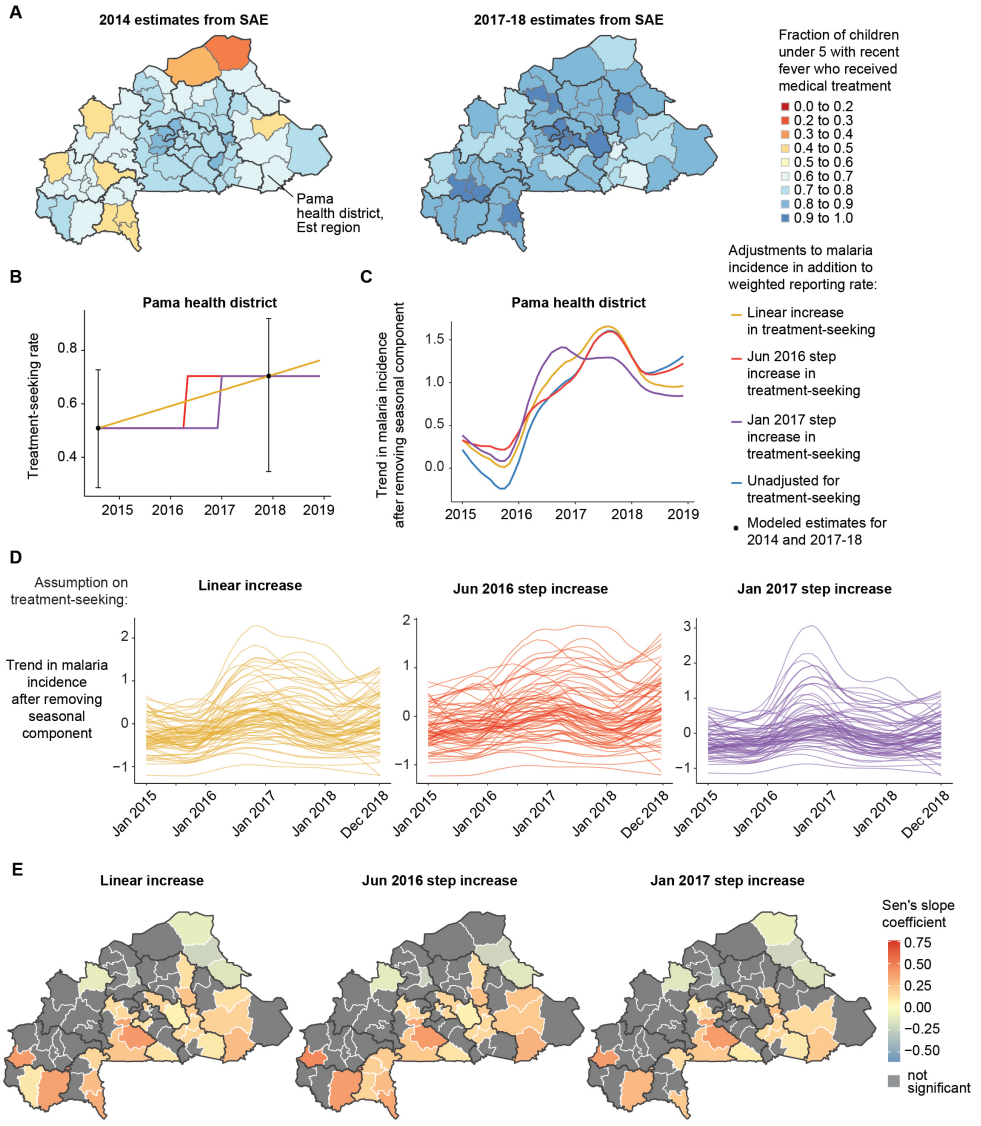


Figure 2.3. Quantifying trends in malaria incidence in children under 5 years old after adjusting for changes in treatment-seeking rates. A) Estimated treatment-seeking rates for recent fevers in children under 5 from 2014 and 2017-18 MIS. B) Three candidate functions used to estimate monthly treatment seeking for districts. Example shown for Pama health district. Black dots and lines indicate the SAE estimates with 95% confidence intervals. C) Trend components of malaria incidence in Pama adjusted for weighted reporting rate alone and malaria incidence adjusted for both weighted reporting rate and each of the three estimates of treatment-seeking. D) Trend components of malaria incidence adjusted for weighted reporting rate and each of the three estimates of treatment-seeking. Each line represents one health district. E) Sens slope coefficient for the trend of malaria incidence adjusted for weighted reporting rate and each of the three estimates of treatment-seeking. Positive numbers indicate increasing trend. Gray: no significant increasing or decreasing trend according to the Mann-Kendall test.

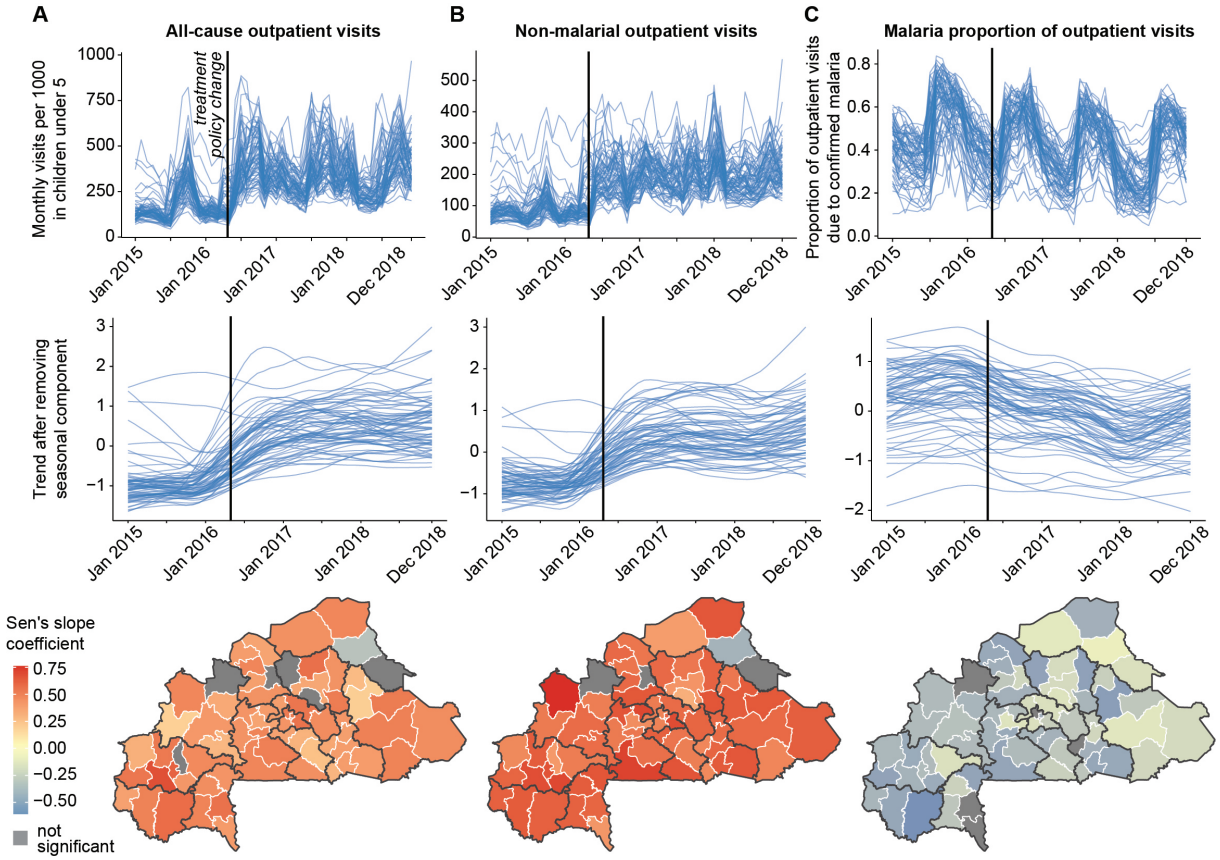


Figure 2.4. Trends in all-cause outpatient visits, non-malarial outpatient visits, and the malaria proportion of outpatient visits in children under 5 years old. A) Monthly all-cause visits to outpatient facilities per 1000 in children under 5 years, monthly non-malarial visits to outpatient facilities per 1000 children under 5 years, and monthly malaria proportion of outpatient visits among children under 5 years old. B) District-level trend components of all-cause outpatient visits, non-malarial outpatient visits, and malaria proportion of outpatient visits. Each line represents one health district, and the black line indicates the 2016 policy change when treatment for childhood illnesses at public health facilities became free of charge. C) Sen's slope coefficient for the trend of all-cause outpatient visits, non-malarial outpatient visits, and malaria proportion of outpatient visits. Positive numbers indicate increasing trend. Gray: no significant increasing or decreasing trend according to the Mann-Kendall test.

2.4. Discussion

Routine case data, with high temporal and spatial resolution, is potentially a very useful data source to inform the strategic planning, monitoring, and evaluation activities carried out by national malaria control programs. Fluctuations in health facility reporting, including both sporadic, incomplete reporting

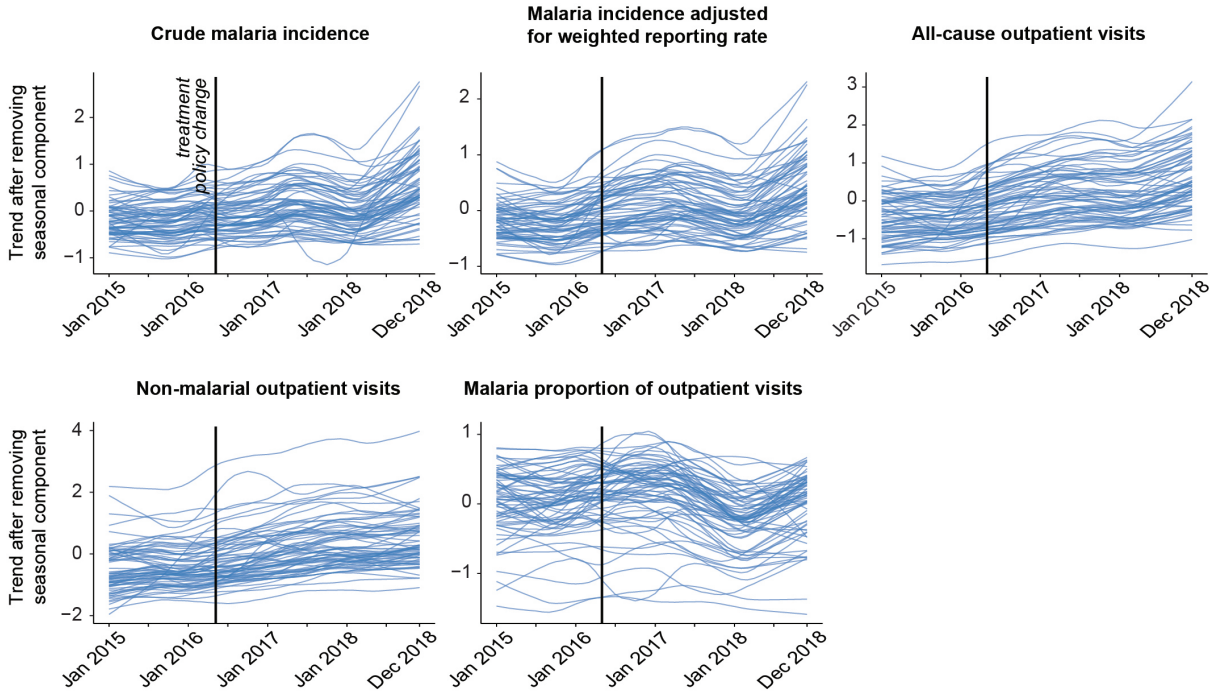


Figure 2.5. Trend components from seasonal trend decomposition for 5 key indicator variables for individuals over 5 years old. Each line represents one health district, and the black line indicates the 2016 policy change when treatment for childhood illnesses in children under 5 years old and for pregnant women at public health facilities became free of charge.

and long-term expansion or improvement of HMIS, can affect a country's capacity to accurately monitor malaria incidence and intervention effectiveness through time. Imputation cannot reliably address issues of many consecutive months of non-reporting. Adjusting reported malaria incidence by additional key factors such as treatment-seeking can be limited by incomplete temporal, spatial, and demographic coverage of the adjustment variables.

Our analysis on the long-term trends in the malaria burden was limited by the relatively short time span of the HMIS data, as 4 years of data is often not enough to employ model-based approaches for trend analysis. Furthermore, Sen's slope can be sensitive to outliers in the time series such as unexpected spikes in incidence. Evaluating consistency between all-cause outpatient visits and confirmed malaria cases has limitations, as reports that have confirmed malaria cases being close to 100% of all-cause outpatient visits may not be realistic for some districts depending on the season. In Burkina, only 0.16% of district-level

observations and 3.4% of health facility-level observations had malaria cases exceed 80% of all-cause outpatient visits. We did not correct for presumed cases and changes in testing practices, as presumed cases were not recorded for all of 2015 and we chose to preserve all 4 years of data. Incorporating presumed cases could potentially flatten the observed trends in malaria incidence because testing increased during this period (Figure A.5) and test positivity rate decreased (Figure A.6). We were not able to include data collected from community health workers. Field studies have found that in Burkina, community health workers do not account for a large source of treatment in areas with health facilities nearby (Druetz et al., 2015; Tipke et al., 2009). However, the lack of case data from community health workers would indicate that our trend estimates may be underestimates.

Improvements in reporting quality between 2015 and 2018 contributed only a small amount to the observed increase in malaria incidence in Burkina Faso. Our methodology to determine active health facilities is limited by the absence of a master list of new health facilities and health facility closures. The availability of information on openings and closures would greatly benefit analysis of routine data, as it could capture temporary health facility closures occurring in the middle of the analysis period. In Burkina Faso, incomplete reporting affected more urban regions and smaller health facilities the most. Reporting rates could have a larger effect in countries or areas where reporting quality experienced larger changes and correcting crude incidence for changes in reporting should be considered when possible.

After adjusting for treatment-seeking, around one-third of districts still exhibited an increasing trend, and very few districts showed a decreasing trend. Changes in treatment-seeking practices among children under the age of 5 between 2015 and 2018 thus could have contributed to the observed increase in reported malaria incidence, but the lack of a clear decreasing trend despite intense intervention is concerning. Treatment-seeking remains difficult to estimate and fine-scale spatio-temporal data is lacking. The MIS before and after the 2016 policy change were conducted in different seasons: the wet season in 2014 and dry season in 2017-18. Treatment-seeking practices may vary seasonally with accessibility of health facilities and parental perception of the need for medical treatment (Ewing et al., 2011; Malik et al., 2006), which may confound the direct comparison of the 2014 and 2017-18 measurements. The SAE for the treatment-seeking rates did not include a temporal correlation component because of the long gap

between survey years and the change in healthcare policy between survey years. It is also possible that treatment-seeking changed further after the 2017-18 MIS survey. The lack of sufficient temporal data on treatment-seeking is a severely limiting factor on the interpretability of routine case data. Furthermore, comprehensive data on treatment-seeking in individuals over 5 is severely lacking.

Trends in incidence adjusted for treatment seeking were flat or decreasing in northern Burkina Faso but still increasing in southern Burkina Faso. Dori and Sebba in the Sahel region had their user fees at health facilities removed for children under 5 years old in September 2008 as part of a pilot trial, prior to the 2016 national change in policy (Zombré et al., 2017). This is reflected in flatter trends in reported malaria incidence and outpatient visits in these two districts (Figure A.7).

For children under 5, trends in confirmed malaria cases, all-cause outpatient visits, and non-malarial outpatient visits all increased between 2015 and 2018. Each showed the steepest increase in early 2016 coincident with the policy change to free treatment for childhood illness, and non-malarial outpatient visits increased the most. This steep increase was not observed for individuals over 5, who experienced a milder increase during the study period. This suggests that while the policy change led to increased treatment-seeking for malaria, treatment-seeking for other illness may have increased even more.

The malaria proportion of outpatient visits among children under 5 years old decreased during the study period for all districts with a significant direction in trend. Other studies of routine case data from Burkina Faso also found that the annual increase in malaria incidence among all ages is not present after accounting for outpatient visits (Rouamba et al., 2020). A decline in the malaria proportion of outpatient visits may be driven by a decline in malaria cases, an increase in incidence of non-malarial diseases, a disproportionate increase in treatment-seeking rate for non-malarial illnesses, or a combination of multiple factors. Since all RDT-positive fevers were recorded as confirmed malaria cases and testing also increased during the study period (Figure A.5), the faster growth of non-malarial outpatient visits could reflect a decline in malaria transmission if treatment-seeking rate was constant. However, little is systematically known about treatment-seeking rate for non-febrile illness, which limits the interpretation of the decreasing trend in malaria proportion of outpatient visits. All adjusted malaria incidence trends

and the trend in malaria proportion of outpatient visits showed an increase in 2018, suggesting that the higher case burden in 2018 may be due to increase in transmission.

The Global Technical Strategy for Malaria 2016-2030 calls for a transformation of malaria surveillance into a core intervention, with routine case data being a key area for investment (World Health Organization, 2015). Limitations in data quality inhibit our ability to understand long term trends in malaria incidence from routine surveillance data. Data completeness and consistency in the future can be improved through training and follow-up on data management guidelines, but understanding existing data is also essential for interpreting trends over many years. To appropriately use routine case data in high-burden countries, it is essential to have better understanding of treatment-seeking patterns, including for non-febrile and non-malarial illness, and a high-quality and robust data collection and reporting process. Approaches such as the methodology presented here enable more rigorous assessment of interventions such as SMC that were rolled out in parallel to major improvements in surveillance quality.

2.5. Conclusion

Improvements in reporting rate and increases in treatment-seeking rates in children under 5 made a small contribution to the increase in malaria incidence observed in routine case data in Burkina Faso between 2015 and 2018. To support the design of passive malaria surveillance and evidence-based decision making with routine case data, it is essential to adjust for sources of bias and translate this data to consistent and reliable measurements.

CHAPTER 3

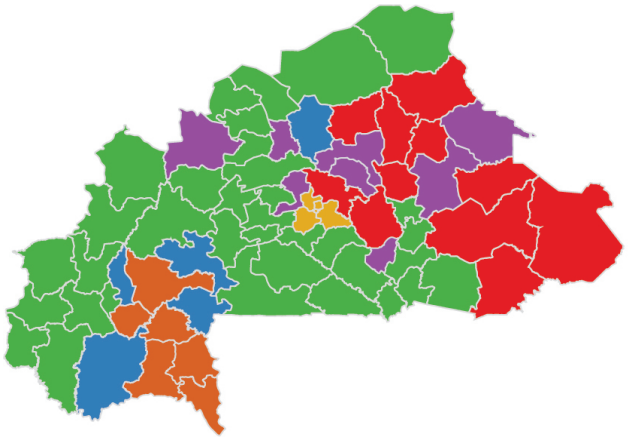
Using routine case data to estimate the protective effectiveness of seasonal malaria chemoprevention under programmatic implementation: a quasi-experimental study in Burkina Faso

3.1. Background

Seasonal malaria chemoprevention (SMC) is implemented in high burden countries in the Sahel region of Africa to prevent uncomplicated and severe malaria in children under 5 years old, who are at highest risk of morbidity and mortality ("Fact Sheet about Malaria," 2022; World Health Organization, 2021). First recommended by the WHO in 2012, SMC is given once a month for 4 months during the peak malaria transmission season (World Health Organization, 2021; World Health Organization and others, 2018b). In 2020, 33.5 million children received at least one dose of SMC in 13 countries in the Sahel region (World Health Organization, 2021).

Meta-analyses of clinical trials have estimated a decrease of at least 75% in clinical malaria with SMC (Cairns et al., 2021; Meremikwu et al., 2012; Wilson & Taskforce, 2011). Clinical trial outcomes do not necessarily represent the real programmatic effectiveness of interventions, making the analysis of SMC effectiveness under programmatic deployment a vital component of understanding the impact of these campaigns. The effectiveness of SMC against uncomplicated malaria under programmatic implementation has been estimated at 15-45% (Baba et al., 2020; Konaté et al., 2020; Rouamba et al., 2020; Wagman et al., 2020), although many of these studies have focused on a limited geographic scope or have analyzed the effectiveness at a broad, national level. The programmatic effectiveness of SMC at a wide, sub-national level remains challenging to estimate in most countries due to limitations in cross-sectional surveys and potentially poor data quality from health facility routine case data.

Burkina Faso is a high-burden country with an estimated 366 confirmed cases per 1000 population at risk in 2020 (World Health Organization, 2021). Most infections occur during or slightly after the peak of the rainy season, approximately between July and November. Children under 5 years old bear the greatest burden, accounting for 43% of all malaria cases and 72% of all deaths attributable to malaria in 2020 (Ministere de la Sante du Burkina Faso, 2019). In 2014 the Burkina Faso National Malaria Control Program (PNLP) started to implement SMC, with all 70 health districts receiving SMC by 2019 (Figure 3.1). Other malaria control campaigns were rolled out around this time, which may have acted in conjunction to SMC in reducing the malaria burden in the country. Long-lasting insecticide-treated bednets (LLINs) were deployed as part of mass distribution campaigns every three years beginning in 2010. Indoor residual sprays (IRS) had also been deployed in certain districts, with additional intervention in 3 districts in 2017 and 2019 (The PMI VectorLink Project, Abt Associates Inc., 2020). Additionally, to fill gaps in access to healthcare for the most vulnerable population groups, treatment at public health facilities became free of charge for children under 5 years old and pregnant women in April 2016 (Ridde & Yaméogo, 2018).



Year of SMC rollout	New SMC districts	Total SMC districts
2014	8	8
2015	11	19
2016	35	54
2017	5	59
2018	6	65
2019	5	70

Figure 3.1. Seventy health districts of Burkina Faso colored by first year of SMC distribution. Districts are labeled by name in Figure B.1.

To assess the programmatic impact of SMC campaigns on a national scale, datasets with long time-trends, high temporal resolution, and large spatial coverage are needed. Routine case data from health facilities can provide a nationally representative view of malaria incidence (Alegana et al., 2020; Tiono

et al., 2014) and has been used to evaluate the programmatic impact of interventions such as rollout of artemisinin-based combination therapy (ACT), LLINs, and IRS in Zanzibar (Ashton et al., 2019) vector control campaigns in Zambia (Chanda et al., 2012); and SMC in Mali (Wagman et al., 2020), Burkina Faso (Kirakoya-Samadoulougou et al., 2022; Rouamba et al., 2020), and across western and central Africa (Baba et al., 2020). One previous study used quality-assured, local routine case data in Mali to assess the effectiveness of SMC, finding a mild reduction of 15% in childhood malaria when compared to districts with no SMC (Wagman et al., 2020). This study focused only on the Segou region in Mali, therefore, their analysis does not capture the full distribution campaign of SMC. The ACCESS-SMC study in west and central Africa utilizes routine case data from DHIS2 in several counties to assess SMC impact. In Burkina Faso, the study finds reductions in the childhood malaria burden of 40.6% in 2015 and 48.5% in 2016 (Baba et al., 2020). Another study in Burkina Faso uses routine case data to analyze the effect of SMC on districts initiating the campaign in 2015 versus districts receiving SMC in 2016. This study finds a 69% reduction in uncomplicated malaria incidence compared to the control group. Both of these studies use data from a limited set of health clinics, exclude major regional hospitals, and do not cover the SMC distribution to all health districts in Burkina Faso (Baba et al., 2020; Kirakoya-Samadoulougou et al., 2022).

To understand the impact of programmatic SMC on preventing clinical malaria in Burkina Faso, we estimated the effectiveness of programmatic SMC with a difference-in-differences model that leverages the staggered rollout of SMC across health districts. We modeled the impact of SMC on malaria incidence, the malaria proportion of outpatient visits, and the incidence of malaria hospitalizations among children under 5 years old and individuals over 5 years old.

3.2. Methods

3.2.1. Data sources

Monthly routine case reports from health facilities from the Burkina Faso Health Management Information System (HMIS) were provided by the Burkina Faso National Malaria Control Program (Programme National de Lutte contre le Paludisme, PNLP). Reports from January 2015 to December 2018 included

2,283 unique health facilities from the 51 districts that began SMC rollout in 2016—2019. We defined the study period as the 16 months in which SMC was distributed in Burkina Faso, corresponding to August to November in 2015, and July to October in 2016, 2017, and 2018, since each round of SMC is expected to protect against clinical malaria for 28 days (World Health Organization and others, 2012).

Fields concerning both children under 5 years old (u5) and individuals over 5 years old (ov5) were used in the analysis. Crude malaria cases were defined as cases confirmed by microscopy ($< 0.3\%$) or by rapid diagnostic test (RDT) ($> 99\%$). All-cause outpatient visits were defined as visits to an outpatient facility due to any cause or illness. The number of malaria hospitalizations was defined as the number of patients who were admitted to inpatient wards due to a malaria diagnosis.

Yearly population estimates for each health district were obtained from PNLP. The fraction of the population u5 was 18% nationally ("World Development Indicators," 2021), and this fraction was applied to all districts. Data on SMC rollout campaigns and dates of SMC rounds in each district were obtained from PNLP. SMC rollout groups were defined as groups of districts initiating SMC distribution campaigns in the same year. Monthly precipitation data was derived from the ERA5 dataset (Hersbach et al., 2020), aggregating the data collected on a 30 km grid to the health district level by averaging the observations in each district. Yearly LLIN use rates for each health district were extracted from the modeled raster layer generated by the Malaria Atlas Project (MAP) (Bertozzi-Villa et al., 2021), where pixel-level data within each district was aggregated by taking the population weighted mean value.

3.2.2. Imputation and data adjustments

We imputed missing all-cause outpatient visits, RDT-confirmed malaria cases, and malaria hospitalizations among children u5 and individuals ov5 at the health facility level. We only imputed observations from time series containing no more than 6 missing observations and no more than 2 consecutive missing values with the seasonal moving average algorithm implemented in the **imputeTS** R package (Moritz & Bartz-Beielstein, 2017). Imputation recovered 2.9% of missing values for all-cause outpatient visits among

children u5 and individuals ov5, 4.1% for RDT-confirmed malaria cases among children u5 and individuals ov5, and 0.8% and 1.2% of malaria hospitalizations for children u5 and individuals ov5, respectively (Table B.1).

We removed all health facilities which had missing data for confirmed malaria cases, missing data for all-cause outpatient visits, or if the number of all-cause outpatient visits was equal to 0 for either age group among children u5 or individuals ov5 for any of the 16 months in the study period, resulting in the removal of 1094/2283 health facilities. Health facilities where the number of confirmed malaria cases was greater than or equal to all-cause outpatient visits in either age group at any point during the study period were also removed (399/2283 health facilities) (Table B.2). We then aggregated data from the remaining 790/2283 health facilities to the health district level.

Data on malaria hospitalizations was more poorly recorded than the other data (Table B.1), and additional health facilities were removed from the dataset when modeling the incidence of malaria hospitalizations. We removed all health facilities which had missing data for the number of malaria hospitalizations among children u5 or individuals ov5 during the study period prior to aggregation at the health district level (Table B.2), resulting in the removal of 2080/2283 health facilities, leaving 203/2283 facilities remaining.

Malaria cases, malaria hospitalizations, and all-cause visits were each adjusted for changes in health facility reporting rate using weighted health facility reporting rates following the methodology in the World Malaria Report (World Health Organization, 2021) and in Chapter 2. We refer to malaria cases per 1000 population at risk after adjusting for reporting rate as malaria incidence. Malaria cases among children u5 were further adjusted for changes in treatment-seeking rates according to u5 treatment-seeking rates estimated using integrated nested Laplacian approximation (INLA) (Lindgren & Rue, 2015; Rue et al., 2009) on Malaria Indicator Survey (MIS) data from 2014 and 2017-18 and a population-weighted district mean (Section 2.2.4), with a step change at June 2016 to account for the 2016 change in healthcare policy (Ridde & Yaméogo, 2018).

The monthly malaria proportion of outpatient visits was calculated by dividing the number of confirmed malaria cases by the number of all-cause outpatient visits, each adjusted for their reporting rates.

3.2.3. Modeling SMC effectiveness

A difference-in-differences modeling approach was applied to the data to quantify the effectiveness of the SMC campaign in Burkina Faso (Huntington-Klein, 2021). The effectiveness of SMC was evaluated by modeling changes in the following indicators among children u5 and individuals ov5 independently: 1) malaria incidence, 2) u5 malaria incidence adjusted for treatment-seeking rates, 3) the malaria proportion of outpatient visits, 4) the incidence of malaria hospitalizations. For each year of SMC rollout, we compared the districts that received SMC for the first time against the districts that had not yet received SMC (Callaway & Sant'Anna, 2021; Huntington-Klein, 2021).

The effect of SMC is estimated using two modeling frameworks. Table 3.1 shows the treatment and control groups used to evaluate the effect of SMC on the 3 different malaria endpoints for the u5 and ov5 age groups. The last column shows the model used. In the first modeling framework (model 1), we use data from only one age group in the study design, assigning individuals living in districts introducing SMC in the same year to unique treatment groups, while the control group consists of individuals living in districts which have not started their SMC campaigns. The second modeling framework (model 2) is reserved only to evaluate the effect of SMC on children u5, where the treatment group consists of children u5 and the control group consists of children u5 in addition to individuals ov5 throughout the country as the latter group never receives the SMC intervention (Table 3.1).

A generalized linear mixed-effects model (GLMM) (McCullagh & Nelder, 1989) was used to quantify the effectiveness of SMC via the incidence rate ratio (IRR), which explains the relative reduction in the malaria indicator in analysis among districts in the SMC treatment group relative to the districts in the untreated control group, and the protective effectiveness ($\text{PE} = (1 - \exp(\text{IRR})) \times 100$. The analysis was restricted to the 4 months during which SMC was administered each year.

The fixed effects model parameters were specific to the difference-in-differences modeling approach and were used to quantify the effectiveness of SMC. The remaining exogenous parameters corresponding to the LLIN use and precipitation were assumed to be random effects varying across districts.

Malaria indicator analyzed	Age group in treatment group	Age group(s) in control group	Model used
Malaria incidence adjusted for reporting rate	u5	u5	model 1
	ov5	ov5	model 1
	u5	u5 & ov5	model 2
Malaria incidence adjusted for reporting rate and treatment-seeking rates	u5	u5	model 1
Malaria proportion of outpatient visits	u5	u5	model 1
	ov5	ov5	model 1
	u5	u5 & ov5	model 2
Incidence of malaria hospitalizations	u5	u5	model 1
	ov5	ov5	model 1

Table 3.1. Malaria indicator analyzed with the corresponding age group, dataset, and model used in the analysis.

For model 1, the model specification is:

$$\begin{aligned}\ln[\mathrm{E}(Y_{t,s,g})] =& \beta_0 + \beta_{1,t^*} + \beta_{2,g} + \delta_g \mathrm{DiD}_{t,g} + \\ & \gamma_{1,s} \ln(1 + X_{t-1,s}) + \gamma_{2,s} \ln(1 + X_{t-2,s}) + \gamma_{3,s} Z_{t^*,s}\end{aligned} \tag{3.2.1}$$

The model specification for model 2:

$$\begin{aligned}\ln[\mathrm{E}(Y_{t,s,g,k})] =& \beta_0 + \beta_{1,t^*} + \beta_{2,g,k} + \beta_{3,k} + \delta_g \mathrm{DiD}_{t,g,k} + \\ & \gamma_{1,s} \ln(1 + X_{t-1,s}) + \gamma_{2,s} \ln(1 + X_{t-2,s}) + \gamma_{3,s} Z_{t^*,s}\end{aligned} \tag{3.2.2}$$

In model 1, $Y_{t,s,g}$ represents either malaria incidence, the malaria proportion of outpatient visits, or the incidence of malaria hospitalizations for date t, year t^*, district s, and initial SMC rollout group g. In model 2, $Y_{t,s,g,k}$ represents either malaria incidence or the malaria proportion of outpatient visits among people in age group k for date t, district s, and initial SMC rollout group g.

The intercept term β_0 is a standard, fixed intercept term. The fixed effect term β_{1,t^*} is the effect of each year t^* to which the date t corresponds. The fixed effect $\beta_{2,g}$ is a group-specific term for each SMC rollout group g from 2016 to 2018. The covariate $\mathrm{DiD}_{t,g}$ is an indicator variable denoting the dates

t of SMC intervention for districts in group g, with $\text{DiD}_{t,g}$ equal to 0 before the initial date t of SMC rollout and 1 after. The regression coefficient of $\text{DiD}_{t,g}$ is a fixed effect denoted as δ_g, which is the IRR on the log-scale. To account for the different age groups incorporated in model 2, we include a fixed effect term $\beta_{3,k}$ which controls for age group differences among children u5 and individuals ov5. Similarly, $\beta_{2,g}$ becomes $\beta_{2,g,k}$ and $\text{DiD}_{t,g}$ becomes $\text{DiD}_{t,g,k}$ to account for individuals ov5 not getting SMC treatment and thus not being counted as part of the treatment group.

The covariates $X_{t-1,s}$ and $X_{t-2,s}$ are the precipitation of the previous month and 2 months prior to date t, with respective random district-specific coefficients $\gamma_{1,s}$ and $\gamma_{2,s}$ that are varying across each district s. The covariates $Z_{t^*,s}$ are the percent LLIN use in year t^*, with the respective random coefficient $\gamma_{3,s}$ that is varying across each district s.

The model accounts for variability in climate of districts in the random coefficients of both lagged precipitation terms ($\gamma_{1,s}$ and $\gamma_{2,s}$). We also account for heterogeneous LLIN use among districts with the random coefficient $\gamma_{3,s}$. The random coefficients are the sum of a fixed effect term γ_j and a random effect $u_{j,s}$ ($j = \{1, 2, 3\}$), where j refers to the random coefficient, and s refers to the district. The random effect of LLIN use is assumed to be normally distributed with zero mean and a variance of σ_1^2 and is taken to be independent of the random effect coefficients of the lagged precipitations. The random effects of the lagged precipitations have a multivariate normal distribution with means 0 and a compound symmetry variance-covariance structure (Huntington-Klein, 2021), with common variance σ^2 and a covariance of $\rho\sigma^2$.

All GLMM modeling were performed with R software using **glmmPQL** in the MASS library (Pinheiro et al., 2007; Venables & Ripley, 2013).

3.3. Results

3.3.1. SMC effect on malaria incidence

We use model 1 to estimate the impact on reducing malaria incidence among children u5 and individuals ov5 separately. The model finds a moderate effect between $27.8 - 44.7\%$ on reducing the adjusted malaria incidence among children u5, with the smallest effect size of 27.8% (95% CI: [18.1%, 36.0%]) observed

for the 2018 SMC rollout group (Table 3.2). SMC was not associated with a significant reduction in the malaria incidence among individuals ov5 for the 2016 and 2018 groups but a 12.2% (95% CI: [1.6%, 21.6%]) PE was associated with SMC for the 2017 group.

We use model 2 and include the ov5 population in the control group to estimate the effect of SMC on the malaria burden among children u5. We find a lower effect of SMC between $8.6 - 34.4\%$ associated with the introduction of SMC in reducing malaria incidence among children u5 compared to using only children u5 in the control group (Table 3.2). The largest difference is seen in the 2016 rollout group, where the PE drops from 38.3% (95% CI: [32.3%, 43.8%]) when excluding individuals ov5 from the control group to 8.5% (95% CI: [2.5%, 14.4%]) with model 2.

After adjusting for changes in treatment seeking behavior, model 1 found a weaker effect of SMC on malaria incidence in children u5 in the 2016 group, where the effect dropped by roughly 11% to an estimated protective effect of 21.7% (95% CI: [19.8%, 33.7%]). The model finds the effect of SMC on the 2017 and 2018 groups to be roughly unchanged after adjusting malaria incidence for changes in treatment seeking behavior (Table 3.2).

3.3.2. SMC effect on the malaria proportion of outpatient visits

Model 1 finds a lower effect of SMC on the malaria proportion of outpatient visits among all groups relative to the analysis on malaria incidence. SMC is associated with a 7.5% (95% CI: [2.3%, 12.5%]) and a 13.7% (95% CI: [7.4%, 19.5%]) reduction in the malaria proportion of outpatient visits in the 2016 and 2017 group, respectively. Furthermore, the model found no significant reduction in the malaria proportion of outpatient visits among districts in the 2018 SMC rollout group associated with the introduction of SMC (Table 3.2). Model diagnostics for the analysis on the u5 malaria proportion of outpatient visits can be found in the supplemental material and show appropriate model behavior for the distributional assumptions (Figure B.3 – B.7). Model 1 also found no significant PE associated with the introduction of SMC in reducing the proportion of outpatient visits due to malaria among individuals ov5 for any of the SMC rollout groups analyzed (Table 3.2).

Model 2 finds a stronger effect of 17.3% (95% CI: [14.1%, 20.4%]) and 18.4% (95% CI: [12.1%, 24.1%]) associated with the introduction of SMC for the 2016 and 2017 rollout groups, respectively. Again, we find that there is no significant PE associated with the introduction of SMC on reducing the malaria proportion of outpatient visits among children in the 2018 rollout group.

3.3.3. SMC effect on malaria hospitalizations

SMC was associated with a moderate reduction between $19.1 - 39.8\%$ in the incidence of malaria hospitalizations among children u5 in districts in all 3 SMC groups (Table 3.2). There was no significant reduction in the incidence of malaria hospitalizations among individuals ov5 associated with the introduction of SMC in the 2016 and 2018 rollout groups, while SMC was associated with a 15.2% (95% CI: [1.2%, 27.2%]) reduction in the 2017 group (Table 3.2).

Yearly LLIN use rates had an interquartile range of $48.0 - 64.8\%$ among all health districts, with a mean use rate of 55.9%. The fixed effect of LLIN use rates was not found to be significant in any of the other models.

Indicator	SMC group	Treated districts	Untreated districts	PE estimates		
				U5 age group (model 1)	Ov5 age group (model 1)	U5 age group (model 2)
Malaria incidence	2016	36	16	38.3% [32.3%, 43.8%]	2.7% [-5.6%, 10.3%]	8.6% [2.5%, 14.4%]
	2017	5	11	44.7% [37.8%, 50.8%]	12.2% [1.6%, 21.6%]	34.4% [25.8%, 42.0%]
	2018	6	5*	27.6% [18.1%, 36.0%]	9.3% [-3.3%, 20.3%]	21.6% [10.6%, 31.3%]
Malaria incidence adjusted for treatment-seeking rate	2016	36	16	27.1% [19.8%, 33.7%]		
	2017	5	11	43.4% [36.2%, 49.9%]		
	2018	6	5*	29.7% [20.2%, 38.1%]		
Malaria proportion of outpatient visits	2016	36	16	7.5% [2.3%, 12.5%]	-4.4% [-10.5%, 1.4%]	17.3% [14.1%, 20.4%]
	2017	5	11	13.7% [7.4%, 19.5%]	0.4% [-7.9%, 8.0%]	18.4% [12.1%, 24.1%]
	2018	6	5*	3.7% [-3.7%, 10.5%]	2.5% [-6.9%, 11.1%]	7.4% [-0.3%, 14.5%]
Malaria hospitalizations	2016	28	14	39.8% [31.4%, 47.2%]	8.5% [-0.8%, 16.9%]	
	2017	4	10	35.2% [20.5%, 47.1%]	15.2% [1.2%, 27.2%]	
	2018	5	5*	19.1% [2.9%, 32.6%]	-6.4% [-23.8%, 8.6%]	

* Districts from Centre region.

Brackets indicate 95% CI.

Table 3.2. Protective effectiveness (PE) estimates on U5 and Ov5 malaria indicators modeled using data at the health district level.

3.4. Discussion

SMC was effective at reducing the malaria burden among children in districts receiving the intervention relative to those which were not receiving the intervention. The difference-in-differences analysis found a moderate effect in reducing the incidence of uncomplicated malaria and malaria hospitalizations among children. Adjusting for changes in treatment-seeking behavior reduced the estimated PE of SMC among districts in the 2016 rollout group. This could indicate a confounding effect of treatment-seeking behavior on changes in reported malaria incidence among children, however, we are limited in fully exploring this effect since we do not have treatment-seeking rates at a monthly temporal scale. SMC was associated with a milder reduction in the malaria proportion of outpatient visits among children in the 2016 and 2017 SMC groups and was not associated with a reduction among children living in districts initiating their SMC campaigns in 2018.

The temporal window of evaluation of this analysis was limited to the period for which electronic data were available (2015 to 2018), meaning we could not analyze the effect on the 2014, 2015, or 2019 rollout groups. Beginning in 2019, all districts in Burkina Faso received SMC, however, a data strike took place during this year, preventing any data routine collection.

A limitation of this quasi-experiment is that the assignment of districts to SMC rollout groups is not random, as the last districts to initiate SMC campaigns were the most urban districts in the Center region of the country, which have a lower incidence of malaria (Soma et al., 2021). This could bias our results as some SMC groups, namely the 2019 SMC group, are not comparable to each other in terms of transmission prior to SMC rollout (Baker et al., 2022; Huntington-Klein, 2021). Analysis on the 2018 rollout group would be the most affected, as this only has the 2019 group for comparison in the control group. Furthermore, the sequential rollout of SMC necessitates careful organization of the difference-in-differences model to evaluate the effect of SMC. Difference-in-differences modeling with multiple time periods is an active area of study, and new methodology which may be more appropriate for this quasi-experimental design may emerge (Baker et al., 2022; Huntington-Klein, 2021).

Our estimated PE of $27.1 - 43.4\%$ in reducing in malaria incidence adjusted for reporting rate and treatment-seeking rates are comparable to those from the ACCESS-SMC observational study of 40.6%

in 2015 and 48.5% in 2016, while both results are lower than the 75% reduction observed in clinical trials and the 88% protective effectiveness found in recent meta-analysis (Baba et al., 2020; Cairns et al., 2021; Meremikwu et al., 2012; Wilson & Taskforce, 2011). Our results also fall below the estimated PE of 69% found in 8 selected health districts starting SMC in 2015 in Burkina Faso (Kirakoya-Samadoulougou et al., 2022), though this study excluded districts with large regional hospitals. The lower effect size be driven by several factors, such as gaps in access to healthcare services, widespread use of RDTs biasing results by over-diagnosing uncomplicated malaria cases (Dalrymple et al., 2018), or confounding effects in reported malaria cases due to increases in treatment-seeking (Rouamba et al., 2020). Our findings show a higher impact than the 15% reduction under programmatic implementation in Mali, which was also based on routine data using RDT confirmed cases (Wagman et al., 2020).

The estimated effectiveness of SMC in reducing the malaria proportion of outpatient visits was much lower than the reductions in malaria incidence found in this study and in the literature, with only the 13.7% reduction being comparable to the 15% reduction observed in Mali. SMC was also not associated with a significant reduction in the malaria proportion of outpatient visits in the 2018 SMC group relative to the 2019 group. This may be driven by the 2019 SMC group being a poor comparison group, however, it could also point to a weak effect of SMC on the 2018 group which were not detected in the analyses of malaria incidence.

SMC was not expected to have a positive effect in reducing the malaria burden among individuals ov5 as the treatment was not intended to be distributed to this group. However, the model found that SMC was associated with a mild reduction of 12.2% and 15.2% in malaria incidence and malaria hospitalizations, respectively, among individuals ov5 in districts in the 2017 rollout group. Analysis in high-burden settings finds a small spillover effect of SMC on individuals ov5 when rolled out to children u5, possibly due to unintended coverage (Scott et al., 2018; Selvaraj et al., 2018). Our analysis is limited in its inability to adjust for changes in treatment seeking among this age group, as the MIS does not cover febrile treatment seeking for the ov5 population. This may confound the associated effect of SMC on this population if the 2018 and 2019 rollout groups, which comprised the comparison group when analyzing the 2017 rollout group, had increased treatment-seeking behavior, and thereby increased reporting of

malaria incidence, relative to the 2017 rollout group. However, trends in malaria incidence and all-cause outpatient visits among this age group suggest no major changes in treatment seeking during this period (Figure A.4). The model found no effect associated with the introduction of SMC in reducing the malaria proportion of outpatient visits among individuals ov5 for any of the SMC rollout groups, meaning that the proportion variable may be controlling for district-level differences in non-malarial factors, like changes in treatment seeking, for which we were not able to explicitly adjust when using malaria incidence as a response in the model.

The inclusion of individuals ov5 into the control group in model 2 to calculate SMC impact on children u5 led to an increase in the estimated effectiveness of SMC in reducing the malaria proportion of outpatient visits for the 2016 and 2017 SMC groups but an overall lower effectiveness of SMC in reducing malaria incidence adjusted for health facility reporting rate. The 17.3% reduction in malaria incidence for the 2016 SMC group we estimate with our model is significantly lower than 48.5% reduction found in Burkina Faso the same year in the ACCESS SMC study (Baba et al., 2020). This may be driven by the inclusion of larger regional hospitals in this analysis, which were excluded from the ACCESS SMC study, if the rate of visits for children u5 compared to individuals ov5 was significantly different than that of other types of health facilities. One limitation in this analysis is that the inclusion of the ov5 age group in the control group of this model inhibits our ability to adjust malaria cases in the u5 group for changes in treatment-seeking rates. Since model estimates for children u5 using malaria incidence adjusted for reporting rate for the 2016 SMC group change drastically when including adjustments for treatment seeking, the inability to include this correction in model 2 could bias estimates.

The estimates of SMC effectiveness presented are specific to reducing the malaria burden at health facilities. This means that there could be a divergence from the true nationwide effect because of imperfect surveillance of the community incidence in passive surveillance data driven by gaps in access to health care (Ashton et al., 2017). Studies analyzing prevalence with cross-sectional surveys using RDTs find a minimum of 44% protective effectiveness of SMC in reducing malaria (Diawara et al., 2017; Druetz et al., 2018). Survey studies analyzing prevalence could yield higher PE depending on when prevalence was measured, the diagnostic method used, characteristics of the sites under surveillance, or the transmission

setting observed. One limitation of the widespread use of RDTs at health facilities is that non-malarial fevers accompanied by a positive RDT result due to asymptomatic malaria infection or lingering antigen presence could be recorded as a confirmed malaria case at a health facility (Dalrymple et al., 2018). This could underestimate the effectiveness of SMC as it may have reduced the severity of malaria infections while RDTs are still confirming these as true cases.

Analyzing the malaria proportion of outpatient visits provides an alternative to modeling malaria incidence as it helps control for increases in malaria reporting due to changes in the country's health system by accounting for it in the denominator. If changes in treatment seeking were equal among malarial and non-malarial illnesses, this indicator would remove the influence of treatment seeking from the temporal trend of malaria burden. With this variable, we also do not have to control for an equal number of health facilities per district for each time point since the magnitude of the proportion would not be affected, reducing the amount of data we would have to discard for modeling. One limitation is that the number of outpatient visits is affected by treatment-seeking and interventions targeting other major diseases, which do not equally affect the numerator. A decline in the malaria proportion of outpatient visits may be driven by an increase in burden of non-malarial diseases, or disproportionate increase in treatment-seeking rate for non-malarial illnesses, rather than a decline in malaria cases.

While the modeling methodology employed does not incorporate the complex spatial dynamics at play between the districts, we incorporate some of the spatial heterogeneity through the random effects terms and present the protective effectiveness of SMC using district-level data. While challenging, the unique use of routine case data provided a useful approach to quantifying the effectiveness of programmatic SMC across districts that would not have been possible with typical cross-sectional or household surveys(Alegana et al., 2020; Ashton et al., 2019). The modeling framework presented here provides an additional use-case on how to leverage routine case data for the evaluation of intervention impact in a setting with a rapidly improving health system.

3.5. Conclusion

SMC was found to reduce the malaria burden among children at health facilities through the use of monthly routine case data. Analyzing the protective effect on malaria incidence is challenging due to the confounding effect of changing reporting rates and treatment-seeking behavior potentially biasing results. The malaria proportion of outpatient visits may control for some of these challenges but introduces new limitations by including non-malarial illnesses in the denominator. The true effectiveness of SMC may lie somewhere in the middle of the estimates of protective effectiveness from either of these malaria indicators. More work in other settings and with higher quality data needs to be done to understand the full extent of the differences between the malaria indicators analyzed.

CHAPTER 4

Fine-scale spatio-temporal modeling of the malaria burden to estimate the protective effectiveness of seasonal malaria chemoprevention using routine case data: a quasi-experimental study in Burkina Faso

4.1. Background

The high burden to high impact (HBHI) initiative to combat malaria in countries with the highest disease burdens calls for the strategic use of readily available datasets to assess the progress of national malaria strategic plans (NMSPs) in the subnational tailoring of interventions (Galatas et al., 2021). The subnational tailoring of interventions is crucial to reduce the malaria burden in high-burden settings, as the limited funds going towards malaria control campaigns must be optimized by targeting the most vulnerable areas and populations. Subnational tailoring of malaria control programs requires high-resolution spatial data collected at healthcare facilities or communities to capture heterogeneities in malaria transmission, intervention history, environmental factors, and demographics. Furthermore, to evaluate the effectiveness of these subnational strategies, this high-resolution spatial data must have time series covering multiple years with data collected at the weekly or monthly scale to capture changing trends in the malaria burden among targeted populations. The fine-scale evaluation of malaria interventions helps locate areas where the intervention may lack adherence or support, helping national malaria control programs target vulnerable regions.

Routine case data in the form of the Health Management Information System (HMIS) in Burkina Faso provides a readily available and regularly updated high-resolution spatio-temporal dataset for the subnational stratification of NMSPs, as it collects monthly data from healthcare facilities around the

country. However, data collected at this spatial and temporal granularity often experiences gaps in reporting (see Chapter 2), which may cause problems in the analysis of disease burden or intervention impact if modeled at a sub-regional or sub-district scale (Ashton et al., 2017). Geo-statistical modeling of fine-scale spatio-temporal data can help solve these issues by leveraging the spatial and temporal correlation structures in the data to fill in the gaps in reporting (Banerjee et al., 2003). This modeling approach enables a more accurate evaluation of subnational trends in malaria burden and intervention impact. Geo-statistical models have been used in the context of mapping disease risk and burden through the use of cross-sectional survey data, passively collected routine case data, and data collected from longitudinal cohort studies. Modeling can leverage the spatial dynamics by exploring the neighborhood structure of data aggregated at the district or regional level using areal units or using distance-based metrics in point-referenced data at the health facility or community/cluster level (Banerjee et al., 2003).

Geo-statistical modeling is prevalent in mapping the disease burden and evaluating the impact of intervention campaigns across various diseases. Mapping the malaria burden across the African continent has been done previously using geo-coded survey data from 49 endemic countries to produce maps at the 1×1 km resolution accounting for spatial and temporal correlations in the data (Noor et al., 2014). The Malaria Atlas Project (MAP) has also made use of fine-scale mapping with geo-statistical models to produce smooth surfaces for various malaria indicators and interventions using data collected from multiple sources (Bertozzi-Villa et al., 2021). Additionally, spatio-temporal modeling has been used in the context of dengue (Wangdi et al., 2018), diarrhoeal disease (Wangdi & Clements, 2017), and tuberculosis (Amsalu et al., 2019) to estimate and visualize high-resolution spatio-temporal data. A study in the Democratic Republic of Congo modeled the outbreak of Ebola using a fine-scale INLA approach in which they analyzed the impact of reported violence on the spread of the disease outbreak (Adegboye et al., 2021).

Routine case data has previously been coupled with geo-statistical models to capture spatial heterogeneities in the malaria burden across regions and assess the performance of local intervention packages. A study in Burkina Faso used HMIS data to forecast clinical malaria incidence through 2019 when the country experienced a workers' strike resulting in large-scale interruptions of routine surveillance. This

study used Bayesian spatio-temporal modeling with structured spatial and temporal random effects to estimate weekly malaria incidence in 2019 and capture the impact of removing healthcare utilization fees for children under 5 years old and pregnant women (Rouamba et al., 2020). Another study in Cameroon using routine case data used Bayesian hierarchical modeling to analyze the spatio-temporal characteristics of malaria incidence among children under 5 years old. The authors in this study used an integrated nested Laplacian approximation (INLA) approach to estimate smooth maps of malaria risk in the country (Danwang et al., 2021). Fine-scale estimations of malaria incidence have also been done in Haiti through routine case data, where Bayesian spatio-temporal modeling was used to capture and map the seasonal fluctuations in case incidence at a 1×1 km resolution (Cameron et al., 2021).

In this chapter, we expand on our efforts in Chapter 3 in which we analyze the impact of SMC using HMIS data aggregated at the health district level. Here, we use a geo-statistical model to leverage the lowest level data in the Burkina Faso HMIS, corresponding to individual healthcare facilities. With this model, we map the fine-scale malaria burden in Burkina Faso at a 1×1 km scale and estimate the impact of seasonal malaria chemoprevention (SMC) on the treated sub-groups, corresponding to the group of districts that initiated the SMC campaign during the same year. Our model employs a differences-in-differences framework which includes multiple treatment periods, which is an active area of study (Callaway & Sant'Anna, 2021; Huntington-Klein, 2021). As noted in earlier sections, SMC has been studied in the context of Burkina Faso (Baba et al., 2020; Cairns et al., 2021; Druetz et al., 2018; Kirakoya-Samadoulougou et al., 2022), but no study has examined its impact using data at the health facility level while leveraging the spatial correlation between the known locations of these facilities. We employ geo-statistical models with INLA frameworks to ease computational power and circumvent complex Markov chain Monte Carlo simulation approaches (Gómez-Rubio, 2020; Rue et al., 2009). We include a modeled latent process describing the fine-scale treatment-seeking behavior, which helps capture the spatio-temporal heterogeneity in changes in the reported malaria burden at healthcare facilities. With the inclusion of the latent process, we have introduced a novel approach to the difference-in-differences framework enables us to capture the impact of SMC while considering the lingering effect of changes in treatment seeking on the malaria burden.

4.2. Methods

4.2.1. Data sources

Monthly routine case reports from healthcare facilities from the Burkina Faso HMIS were provided by the Burkina Faso National Malaria Control Program (Programme National de Lutte contre le Paludisme, PNLP). Reports were collected from 2,959 healthcare facilities across all 70 health districts and the data ranged from January 2015 to December 2018. Out of 2,959 healthcare facilities, 1,235 (41.7%) were geographically tagged with GPS coordinates and were retained for modeling while the remaining 1,724 (58.3%) were discarded.

We model the time series of the malaria proportion of outpatient visits, which is defined as the number of confirmed malaria cases over the number of all-cause outpatient visits, among children under 5 years old. Malaria cases are confirmed by microscopy ($< 0.3\%$) or by rapid diagnostic test (RDT) ($> 99\%$). All-cause outpatient visits were defined as visits to an outpatient facility due to any cause or illness. We imputed missing RDT-confirmed malaria cases and all-cause outpatient visits among children under 5 years old according to the same methodology detailed in Chapter 3 (Table B.1).

To model the malaria proportion of outpatient visits at the healthcare facility level, we remove all facilities which have any missing all-cause outpatient or 0 reported all-cause outpatient visits for any of the 48 time points in the study period, leaving 1,190 healthcare facilities remaining (45 discarded). We also aggregate healthcare facilities that are within the same 1×1 km grid square, which aggregates 4 healthcare facilities into 2: one aggregated healthcare facility in the district of Kombissiri and one aggregated healthcare facility in the district of Manga.

Furthermore, we only include healthcare facilities from health districts starting SMC in 2016, 2017, 2018, and 2019 (see Chapter 3, Figure 3.1), leaving 816 healthcare facilities (Figure 4.1). We call the group of districts starting SMC campaigns in the same year an SMC rollout group. Data on SMC rollout campaigns and dates of SMC rounds in each district were obtained from PNLP.

We considered climate covariates for monthly precipitation and monthly surface air temperature in the model. Monthly precipitation and monthly surface air temperature data were derived from the ERA5

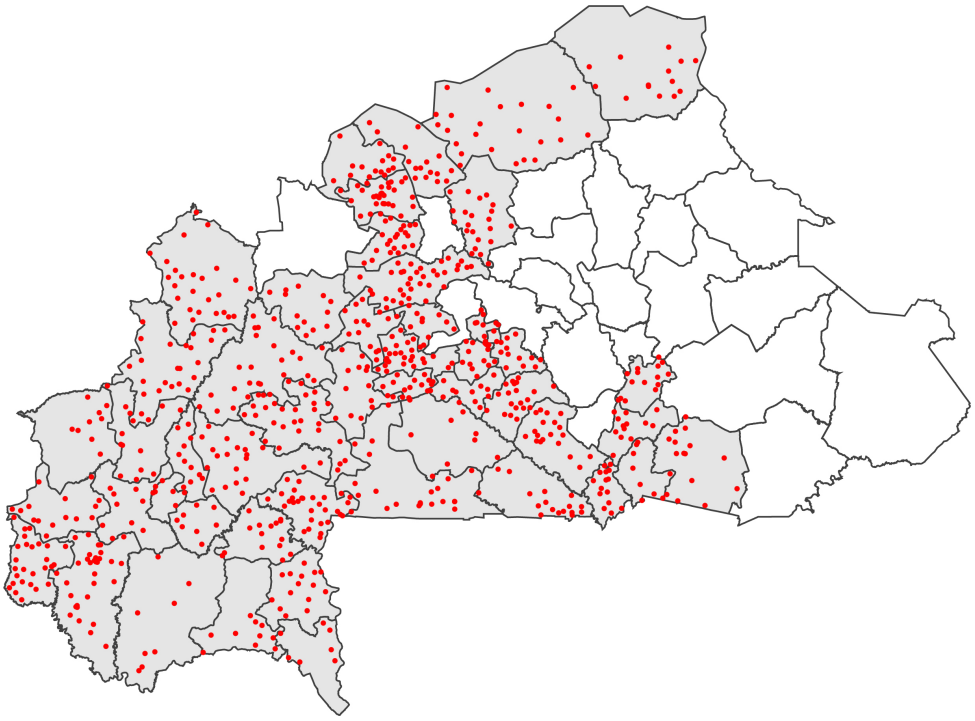


Figure 4.1. Location of the 816 health facilities used for modeling from districts in the 2016, 2017, 2018, and 2019 SMC rollout groups.

dataset, which collects data at a 30 km grid mesh (Hersbach et al., 2020). Climate values were matched to each healthcare facility by taking the average climate values of grid points surrounding each facility. We considered the climate variables for the current month and for the two months previous (lag-1 and lag-2) in the model. Model diagnostics showed that the model did not significantly improve with any of the climate covariates or their lagged values (Appendix C, Table C.1), thus they were omitted from the final model.

We have shown that changes in treatment-seeking behavior influence the trend of the reported malaria burden at healthcare facilities (Chapter 2). Thus, we model the fine-scale treatment-seeking rate at a 1×1 km grid among children under 5 years old using a latent process as described in the previous chapters (see A.2). Treatment-seeking rates are associated with each health facility by averaging the values of the grip points surrounding each health facility.

4.2.2. Model formulation for SMC effectiveness

A difference-in-differences modeling approach is used to estimate the effectiveness of the SMC campaign in Burkina Faso (Huntington-Klein, 2021). Due to the sequential rollout of the SMC campaign, we use multiple treatment periods in the modeling framework to account for the different SMC rollout groups. For each year of SMC rollout, we compared the districts that received SMC for the first time against the districts that had not yet received SMC (Callaway & Sant'Anna, 2021; Huntington-Klein, 2021).

The difference-in-differences model for SMC effectiveness follows a generalized linear mixed-effects modeling (GLMM) approach with an integrated nested Laplacian approximation (INLA) framework in which we capture the spatial heterogenity at the healthcare facility level and the temporal variation with a seasonal effect (Ingebrigtsen et al., 2014; Rue et al., 2009). Using the difference-in-differences approach, we capture the impact of SMC using the incidence rate ratio (IRR), and translate this to a protective effectiveness ($\text{PE} = (1 - \exp(\text{IRR})) * 100\%$). We estimate the effectiveness of SMC by modeling the changes in the malaria proportion of outpatient visits with the aid of a latent process describing the fine-scale treatment-seeking rate which we have shown to influence the estimated effectiveness of SMC (see Chapter 3). Our novel approach to the differences-in-differences framework enables us to capture the impact of SMC and the modeled latent treatment-seeking behavior on the malaria proportion of outpatient visits (Appendix C, Section C.2).

Due to the influence of changes in treatment-seeking behavior on the reported malaria burden, we model it as a latent process with a GLMM INLA model using data from the 2014 and 2017/18 malaria indicator surveys (MIS) (ICF International, 2015, 2019), where we model each survey year separately. We incorporate structured spatial random effects with a Matérn covariance function (Lind-gren & Rue, 2015; Rue et al., 2009), unstructured spatial random effects at the district level, and climate covariates of interest. Further detail on this can be found in Appendix A.2.

For a healthcare facility h, L_h is then the associated treatment-seeking rate, where the fine-scale model is estimated at a 1×1 km resolution from cluster-level observations. We temporally vary L_h by

joining the estimates for the 2014 and 2017/18 survey years with a step-function where the jump is placed in June 2016, which corresponds most closely to the change of healthcare policy. The cluster-level model for the treatment-seeking rate (π_j) for the 2014 and 2017/18 surveys is estimated as follows.

Let Y_{jD} be the number of children who seek medical treatment for a recent fever in cluster j which is located inside health district D. N_{jD} is the total number of children who had a recent fever in that cluster. We assume that Y_{jD} follows a Binomial distribution, $Y_{jD}|N_{jD}, \pi_{jD} \sim \text{Bin}(N_{jD}, \pi_{jD})$, where π_{jD} is the proportion of children who seek medical treatment for a recent fever.

We model the latent process $\eta_{jD} = \text{logit}(\pi_{jD})$ as follows:

$$\begin{aligned} Y_{jD}|\eta_{jD} &\sim \text{Bin}(N_{jD}, \pi_{jD}) \\ \eta_{jD} &= X_j\beta + v_D + u_j \\ \mathbf{v}|\tau_v &\sim \text{N}(0, \Sigma_{\mathbf{v}}) \\ \mathbf{u}|\theta_u &\sim \text{GF}(0, \Sigma_{\mathbf{u}}) \end{aligned} \tag{4.2.1}$$

where X_j is the matrix of covariates for the model at cluster location j, v_D is the random effect associated with the district D to which cluster j belongs with an i.i.d. Gaussian covariance structure, and u_j are structured spatial random effects with a Matérn covariance structure (Rue et al., 2009).

The district random effect $\mathbf{v} = (v_1, \ldots, v_n)$ has a Gaussian distribution with a general i.i.d. covariance structure with hyperparameter τ. The spatial SPDE random effect $\mathbf{u} = (u_1, ..., u_m)$ follows a Gaussian random field process that captures the fine-scale spatial dependence. For the spatial correlation structure $\Sigma_{\mathbf{u}}$, we use a Matérn covariance function, which enables the SPDE specification (Rue et al., 2009). This assumes that the correlation between two clusters j and j' decreases as the distance between the two clusters increases, and vice versa (Ingebrigtsen et al., 2014; Lindgren & Rue, 2015).

Assuming that the treatment-seeking rate follows the latent process described above (Equation 4.2.1), we model the time series of the malaria proportion of outpatient visits at healthcare facilities to estimate the effectiveness of SMC where the linear predictor in the model includes the latent treatment-seeking rate. Climate covariates considered did not significantly improve the model performance, thus were omitted

from the final model (Table C.1). We estimate the changes in the malaria proportion of outpatient visits in the treated districts versus the untreated districts at the healthcare facility level as follows.

The model consists of time series of confirmed malaria cases per health facility $Y_{t,h}$ where t is time in months ($t \in \{1, 2, \ldots, 48\}$) and h corresponds to a unique health facility. $Y_{t,h}$ follows a Poisson distribution with rate $\lambda_{t,h}$ such that $Y_{t,h}|M_{t,h} \sim \text{Pois}(\lambda_{t,h})$, where the rate $\lambda_{t,h}$ is the proportion of confirmed malaria cases over the number of all-cause outpatient visits $Y_{t,h}/M_{t,h}$ at time t and health facility h.

The model for the malaria proportion of outpatient visits, which depends on the latent process for the treatment-seeking rate, is then,

$$\begin{aligned} Y_{t,h}|\lambda_{t,h} &\sim \text{Pois}(\lambda_{t,h}) \\ \ln(\lambda_{t,h}) &= \alpha_{y(t)} + \gamma_{g(h)} + \delta_{t,g(h)} + \beta_L L_{t,h} + \rho_t + \omega_{t,h} \\ (\rho_i + \cdots + \rho_{i+12-1})|\tau_\rho &\sim \text{N}(0, \Sigma_{\boldsymbol{\rho}}) \\ \boldsymbol{\omega}|\theta_\omega &\sim \text{GF}(0, \Sigma_{\boldsymbol{\omega}}) \end{aligned} \tag{4.2.2}$$

where $\alpha_{y(t)}$ is the year fixed effect such that $y(t) \in \{2015, \ldots, 2018\}$, $\gamma_{g(h)}$ is the SMC rollout group fixed effect for rollout group $g(h)$, and $\delta_{t,g(h)}$ is the term of the difference-in-differences estimator attributable to SMC (Appendix C, Section C.2). $\delta_{t,g(h)}$ is an indicator variable equal to 0 before the first time (t) that the SMC group ($g(h)$) initiates the SMC campaign and 1 after. $L_{t,h}$ is a latent process describing the treatment-seeking rate associated with a health facility h at time t with associated coefficient β_L. ρ_t is a seasonal temporal random effect, and $\omega_{t,h}$ is a structured spatio-temporal random effect with a Matérn spatial covariance function (Rue et al., 2009).

The temporal random effect ρ_t describes the seasonal variation in the data with an assigned periodicity of 12 months, describing the yearly seasonal pattern of malaria transmission. The random vector $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_{48})$ assumes that the sums $(\rho_i + \cdots + \rho_{i+m-1})$ are independent Gaussians with periodicity m. The spatio-temporal random effect $\boldsymbol{\omega}$ follows an SPDE specification that is i.i.d. across time, with $\boldsymbol{\omega} = (\omega_{t,1}, \ldots, \omega_{t,k})$ following a Gaussian random field process that captures the fine-scale spatial dependence

which is i.i.d. across the 48 months in the data. For the spatial correlation structure $\Sigma_{\boldsymbol{\omega}}$, we use a Matérn covariance function, which enables the SPDE specification (Rue et al., 2009). This assumes that the correlation between two health facilities h and h' decreases as the distance between the two clusters increases, and vice versa (Ingebrigtsen et al., 2014; Lindgren & Rue, 2015).

The best model was selected with a step-wise selection framework in which a climate covariate was included in the model if it showed to be significant in a model with only that climate covariate considered. The final model was selected based on the smallest values of the deviance information criterion (DIC) and Watanabe–Akaike information criterion (WAIC) (Ingebrigtsen et al., 2014) (Table C.1). Model estimated hyperparameters are shown in Appendix C (Table C.2).

4.2.3. Mapping the estimated proportion

We produce smoothed maps of model estimates for the malaria proportion of outpatient visits among children under 5 years old. Maps are generated using a 1×1 km grid, where covariate values are matched to each grid point with the using the same resolution. The maps are smoothed by fixing the spatio-temporal random effect $\omega_{t,h}$ at its 0 mean value.

Maps are produced for each month of the high transmission season, as these are months in which SMC is being rolled out to the SMC groups. This results in 16 maps, where the months shown are July, August, September, and October from $2015 - 2018$. Maps for the entire 48 months in the study period are presented in Appendix C (Figure C.1).

All modeling was done with R software using the **R-INLA** package (Krainski et al., 2018).

4.3. Results

4.3.1. SMC effect on the malaria proportion of outpatient visits

We find a moderate protective effectiveness of SMC in reducing the malaria proportion of outpatient visits, ranging between $10.9 - 16.2\%$ on all SMC rollout groups analyzed 4.1. The effect is weakest in the 2016 SMC rollout group, with a protective effectiveness of 10.9% (95% CI: [8.3%, 13.5%]). The strongest effect was found in the 2017 SMC rollout group, which had an associated 16.2% PE (95% CI:

[12.9%, 19.3%]). The protective effect on the 2018 group showed a moderate effect of SMC in reducing the malaria proportion of outpatient visits, showing a 11.0% (95% CI: [5.8%, 15.9%]). The PE for the 2018 SMC group also had the widest confidence interval, showing that this group had a larger uncertainty associated with the effectiveness of SMC.

The estimates of protective effectiveness for the 2016 and 2017 SMC groups are comparable, yet slightly stronger, than the PEs found in Chapter 3 (Table 3.2). Using the HMIS data at the healthcare facility level in a spatio-temporal model, we are able to capture the effectiveness in the 2018 SMC group, which showed to be insignificant in the district-level model in Chapter 3 (Table 3.2).

			PE estimates	
	Treated districts (#HFs)	**Untreated districts (#HFs)**	**mean**	**95% CI**
Effect on 2016 SMC rollout group	34 (603)	16 (213)	10.9%	[8.3%, 13.5%]
Effect on 2017 SMC rollout group	5 (101)	11 (112)	16.2%	[12.9%, 19.3%]
Effect on 2018 SMC rollout group	6 (75)	5* (37*)	11.0%	[5.8%, 15.9%]

* Districts or health facilities (HFs) from Centre region.
Brackets indicate 95% CI.

Table 4.1. Protective effectiveness (PE) estimates of SMC on children under 5 modeled using data at the healthcare facility level.

4.3.2. Model coefficient estimates

We show the estimates for the log-scale model coefficient on the covariates in table 4.2 below. We show the estimates for the fixed effects associated with each SMC rollout group and the year fixed effect. We also show the coefficient estimate for the latent treatment-seeking rate. Lastly, we show the coefficient estimate for the difference-in-differences (DiD) estimators for the SMC effectiveness on each rollout group, which has been transformed into a protective effectiveness in table 4.1 above.

	log-scale coefficient estimates	
Model covariates	**mean**	**95% CI**
2016 SMC group effect	-0.683	[-0.704, -0.662]
2017 SMC group effect	-0.617	[-0.648, -0.587]
2018 SMC group effect	-0.591	[-0.623, -0.559]
2019 SMC group effect	-0.941	[-0.981, -0.901]
2016 year effect	-0.049	[-0.082, -0.017]
2017 year effect	-0.195	[-0.234, -0.157]
2018 year effect	-0.321	[-0.361, -0.280]
Latent treatment-seeking rate	0.023	[0.017, 0.030]
2016 SMC DiD effect	-0.116	[-0.145, -0.087]
2017 SMC DiD effect	-0.176	[-0.215, -0.138]
2018 SMC DiD effect	-0.116	[-0.173, -0.060]

Brackets indicate 95% CI.

Table 4.2. Model covariates and log-scale coefficient estimates used for healthcare facility-level model to estimate the malaria burden and capture the effectiveness of SMC.

The fixed effects associated with the SMC rollout groups show that being associated with an SMC rollout group that gets treated within the study period (i.e. the $2016 - 2018$ rollout groups) had a strong effect in reducing the malaria proportion of outpatient visits, while being in the 2019 rollout group had a weaker association with a reduction in the malaria proportion of outpatient visits (Table 4.2).

The latent treatment-seeking rate had a positive coefficient value associated with it, meaning that the higher the treatment-seeking rate, the higher value of the proportion of malaria cases at health facilities is

observed and reported. The log-scale value is small, meaning that this is a relatively weak, yet significant, effect.

4.3.3. Maps of estimated malaria proportion of outpatient visits

Maps of modeled estimates for the malaria proportion of outpatient visits among children under 5 years old are shown in figure 4.2 for the months of June−October. The first row of the figure, which corresponds to the SMC months in 2015, shows modeled estimates during the pre-SMC time period for all healthcare facilities in districts analyzed. Estimates from July 2015 show a relatively low burden of malaria during this month, but we can see the much higher values of the malaria burden during the August through October months. The highest burden appears to be in the southern region of Sud-Ouest as well as some districts in Cascades, Hauts Bassins, and Boucle du Mouhoun. The lowest burden is estimated to be in the Centre region, which is expected (Figure 4.2).

As we move into the SMC months in 2016, we see a large reduction in the malaria proportion of outpatient visits among districts in the 2016 SMC rollout group outlined in red (Figure 4.2). This is most apparent in the months of August through October, though estimates from October show a slight increase in the malaria proportion of outpatient visits among all districts.

Similarly, in 2017 we see large reductions in the estimated malaria proportion of outpatient visits compared to the previous year among districts starting SMC this year which are outlined in red (Figure 4.2). The districts in the 2016 SMC group see a sustained reduction compared to estimates from maps in 2015, with some reductions seen in September and October compared to the estimates from 2016.

Finally, districts in the 2018 SMC group see a slight reduction in the malaria proportion of outpatient visits when they initiate the SMC campaign in July 2018 (Figure 4.2). These districts still have the highest estimated malaria burden throughout the country, though the reductions, especially compared to the estimates of their malaria burden in 2015 and 2016, are substantial.

No large change is seen in the districts in the Centre district, located in the middle of the country, as these districts were not scheduled to receive SMC until the high-transmission season of 2019, which is beyond the study period.

Maps covering all 48 months in the study period are presented in Appendix C (Figure C.1).

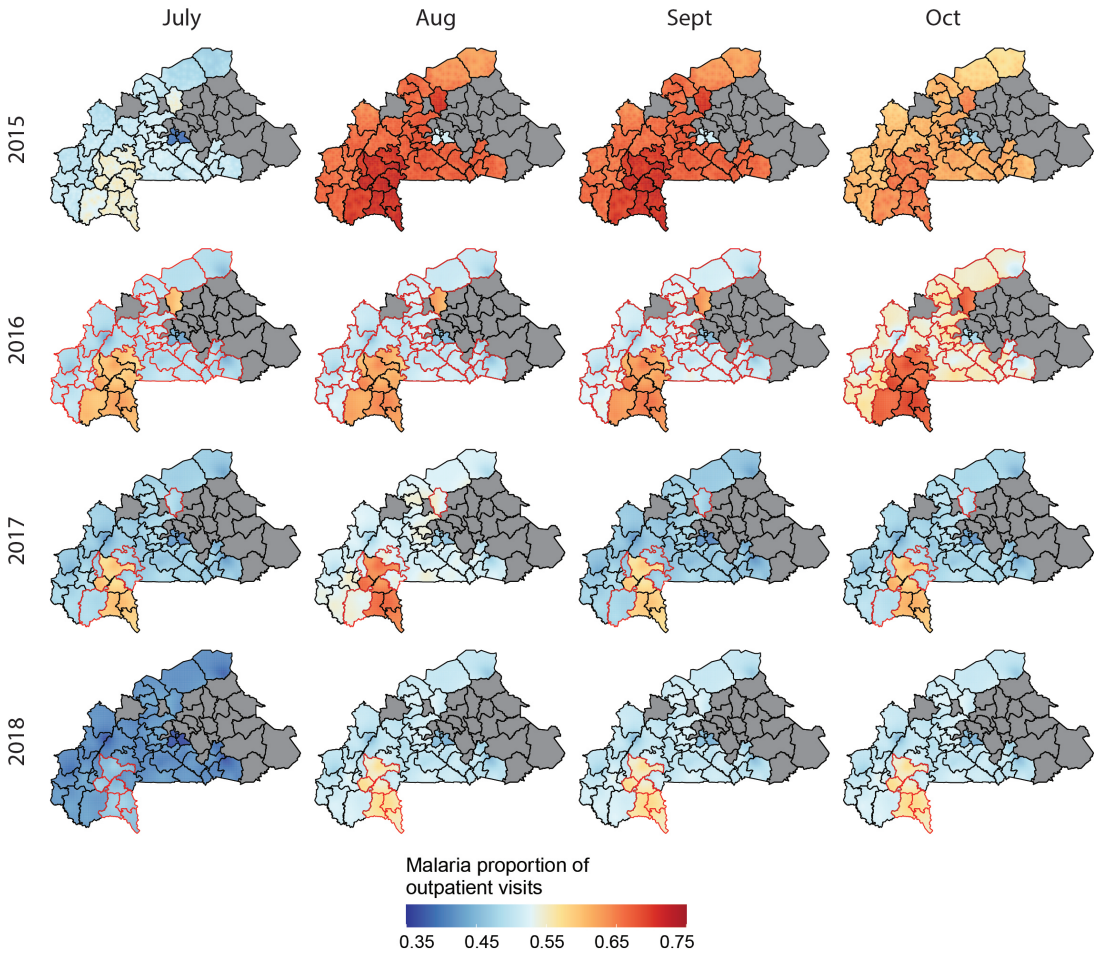


Figure 4.2. Maps of modeled estimates for the malaria proportion of outpatient visits at a 1 x 1 km scale. The months mapped correspond to the months of SMC distribution. Districts are outlined in each map, and red outlines indicate districts which initied SMC in that year. Districts with grey color indicate districts which were not included in the model.

4.4. Discussion

SMC shows modest effectiveness in reducing the malaria proportion of outpatient visits among children under 5 years old in healthcare facilities residing in districts receiving the intervention relative to those not receiving the intervention. The difference-in-differences analysis shows that SMC has a protective effectiveness ranging from $10.9-16.2\%$. This impact is visible in the mapped estimates of the malaria

proportion of outpatient visits in Figure 4.2. One key improvement in this section is that, in contrast to the results in Chapter 3, we find a significant impact of SMC on reducing the malaria proportion of outpatient visits in the 2018 SMC rollout group. The novel approach to the differences-in-differences framework presented in this section enables us to capture the impact of SMC and the modeled latent treatment-seeking behavior on the malaria proportion of outpatient visits (Appendix C, Section C.2). A key limitation in this study is the low percentage of healthcare facilities which were geo-located and included in the model, reducing our ability to fully analyze the effect among all treated health districts and reducing the statistical power the capture the true effect size.

As in Chapter 3, the temporal window of the study period prevented us from analyzing all SMC rollout groups in the programmatic SMC campaign. The lack of data from 2013 and and 2014 limits our ability to analyze the effect on the 2014 SMC group as we have no data for the pre-intervention period for this rollout group. Similarly, analyzing the effectiveness of SMC on the 2015 rollout group is limited since we have no data from a pre-intervention, high-transmission period—corresponding to the months of July through October in 2014—for this rollout group. Finally, we have no ability to analyze the effectiveness on the 2019 rollout group since we have no data in the post-intervention period for this group nor do we have any data for a control group to use for comparison in the difference-in-differences analysis.

Many of the basic limitations in this quasi-experiment carry over from those listed in Chapter 3, namely: the lack of random assignment of districts to treatment groups, the lack of comparability between SMC groups, specifically relative to the 2019 SMC rollout group, and travel to large regional hospitals from outside districts implying a possible contamination of data in healthcare facilities in treated districts from non-treated children. Districts in the Centre region—which see the lowest malaria transmission rates and the lowest malaria burden at a baseline (Soma et al., 2021)—receive SMC has part of the last rollout group in the campaign. This would affect estimates of protective effectiveness for the 2018 SMC rollout group the most, as it only has the 2019 SMC group as a comparison. Additionally, districts in SMC groups receiving the intervention earlier may have been targeted due to their high levels of baseline transmission and burden. Finally, travel to large regional hospitals for treatment may contaminate data. Children from non-treated districts may travel to large regional hospitals in SMC-treated districts, potentially

reducing the estimated impact of SMC in the SMC rollout group which contains those treated districts. The opposite effect may also be true, where children in SMC-treated districts may seek treatment to large hospitals in non-treated districts, which again would reduce the estimated impact of SMC in the SMC-treated districts.

The sequential rollout of the SMC campaign in Burkina Faso necessitates careful organization of the treatment groups in the difference-in-differences model to evaluate the effect of SMC. The framework allows us to compare changes in the malaria burden among healthcare facilities during the time period previous to SMC treatment against the time period after SMC treatment for those districts that recently introduced SMC treatment relative to those districts that have yet to receive SMC treatment. Currently, difference-in-differences modeling frameworks are unable to account for SMC being distributed only during the 4 months of the high-transmission period each year, and must assume that the intervention is persistent once introduced into a treatment group (Huntington-Klein, 2021). This should have a minor effect on the SMC effectiveness, as the yearly trends are dominated by the malaria burden during the 4 months of the high-transmission season in which SMC is distributed since this is the period in which the bulk of the yearly malaria burden occurs. Furthermore, the inclusion of the seasonal temporal random effect helps control the trend for the seasonal fluctuations in the malaria burden. Our approach to the differences-in-differences framework captures the impact of SMC and the modeled latent treatment-seeking behavior on changes to the malaria proportion of outpatient visits. The differences in trends among healthcare facilities in treated versus untreated districts is not solely due to the introduction of SMC, as the modeled latent process for the treatment-seeking rate also influences the estimated changes in the malaria burden before and after the national change in healthcare policy affecting children under 5 years old (Appendix C, Section C.2). The inclusion of this latent process is necessary, as we have shown that the changes in treatment-seeking behavior influence the observed trend of the malaria proportion of outpatient visits (Chapter 2, Figure 2.3) and the impact of the SMC campaign (Chapter 3, Table 3.2). Difference-in-differences analysis with multiple treatment periods is an active area of study, and new developments in literature may arise to address these limitations (Callaway & Sant'Anna, 2021; Huntington-Klein, 2021).

The estimated protective effectiveness of SMC at reducing the malaria proportion of outpatient visits at healthcare facilities was comparable to the estimates on the district-aggregated data from Chapter 3. For the 2016 SMC rollout group, we estimated a PE of 10.9% (95% CI: [8.3%, 13.5%]) in this chapter, while the PE in Chapter 3 was 7.5% (95% CI: [2.3%, 12.5%]). This shows that the point estimates are comparable to each other, as the confidence interval for the result in Chapter 3 contains the 10.9% point estimate from this seciton. For the 2017 SMC group, we estimated the highest PE among all SMC groups, with an estimated PE of 16.2% (95% CI: [12.9%, 19.3%]). The PE estimates for the 2017 SMC group in Chapter 3 was 13.7% (95% CI: [7.4%, 19.5%]), which again shows that the results are comparable and within each other's confidence intervals. For the 2018 SMC group, we estimated a PE of 11.0% (95% CI: [5.8%, 15.9%]), while the effect was not significant in the estimates presented in Chapter 3. The significant PE estimate for this group could be due to the structure of model explored in this section, where we utilize the entire 48 months of the time series versus only 16 months of the high-transmission period. We further leverage the full time series by applying a seasonal random effect to the temporal component, helping us estimate the seasonal fluctuations in the malaria burden. Finally, the structured spatial random effect using the Matérn correlation structure helps borrow information from neighboring health facilities to account for gaps in reporting quality among healthcare facilities residing in districts in the 2018 SMC group and helps account for the small sample sizes in the 2018 SMC group and its comparison group, corresponding to the 2019 SMC group.

We were unable to analyze the protective effectiveness of SMC at reducing the malaria incidence since we have no data for the catchment populations or catchment areas of healthcare facilities. Gravity models have been used in previous studies to estimate catchment areas and populations (Cameron et al., 2021). However, we are limited in our ability to apply some of these modeling based and naïve approaches to estimate catchment populations of healthcare facilities due to the low number of facilities geo-referenced with coordinate locations. Attempts to estimate the population denominator for malaria incidence at the health facility level using data on antenatal care visits or intermittent preventive treatment of malaria for infants or pregnant women were affected by poor reporting in a majority of facilities and inconsistent reporting quality across facilities. If presented with more complete data on the geo-location of health

facilities, this study could be expanded to explore the effect of SMC on reducing the malaria incidence at the health facility level.

The estimates of protective effectiveness for SMC presented in this study are specific to reducing the malaria burden at health facilities. There may be a discrepancy from the true effect in the community due to imperfect surveillance of the burden in passive surveillance data driven by gaps in access to health care and gaps in reporting quality (Ashton et al., 2017).

The malaria proportion of outpatient visits provides an alternative indicator with which to model changes in the malaria burden in the healthcare sector when there are limitations in estimating the malaria incidence. The malaria proportion of outpatient visits helps to control for changes in treatment-seeking behavior at the health facility level by helping to control for increases in reporting in the numerator as well as in the denominator of the calculated proportion. A limitation of this indicator of the malaria burden is that the number of outpatient visits may be affected by treatment-seeking, interventions, and changes in transmission or incidence among other major diseases, which would not affect the numerator equally. A decline in the malaria proportion of outpatient visits may be influenced by increases in the burden of non-malarial diseases or disproportionate increase in treatment seeking for non-malarial illnesses rather than a decline in the trend of malaria cases.

The modeling approach presented in this study shows the impact of analyzing the malaria burden at scale of the healthcare facilities in the Burkina Faso HMIS. In contrast to estimates of SMC effectiveness using district-aggregated data (Chapter 3), this model does capture the effect on the 2018 SMC rollout group, showing the positive impact of the intervention in reducing the malaria burden. Additionally, this methodology allows us map the malaria burden at a granular spatial resolution, in this case a 1×1 km scale. The inclusion of the spatial and temporal structured random effects proved to be influential in the model and helped improve model behavior and estimated results of intervention effectiveness. The inclusion of the latent process for modeled treatment-seeking rate helped control for heterogeneous changes in treatment seeking at the sub-district level. This modeling framework necessitates accurate and complete information for GPS locations of health facilities in the HMIS data, and results presented

could be improved with more complete information on the geo-locations of health facilities not included in the analysis.

4.5. Conclusion

We found SMC to have a positive effect on reducing the malaria burden in health facilities across Burkina Faso through the use of passively collected routine case data. Analyzing the protective effectiveness of interventions through routine case data is challenging due to gaps in reporting quality, and geo-statistical modeling approaches can leverage spatial and temporal correlation structures to alleviate these challenges. Furthermore, including modeled latent processes for factors influencing the reported malaria burden at healthcare facilities improves the accuracy of the estimated malaria burden and estimates of the impact of interventions. The inclusion of the latent process expands on the ongoing work in differences-in-differences modeling with multiple treatment periods by proposing a novel approach to the estimation of the changes in the malaria burden. This methodology may be extended to study other diseases or analyze the malaria burden and the impact of interventions in different settings.

CHAPTER 5

Conclusion and Future Directions

The quasi-experimental studies conducted in Chapters 3 and 4 using the routine case data from Burkina Faso's health management information system (HMIS) showed that seasonal malaria chemoprevention (SMC) was associated with a mild to moderate protective effectiveness against symptomatic malaria among children under 5 years old. Using routine case data poses challenges, as changes in treatment-seeking behavior and health facility reporting quality partially contributed to the reported increase in the malaria burden between 2015 and 2018 observed in Chapter 2. Applying adjustments for these factors did not completely remove the increasing trend in the data. This is partially because of our inability to fully estimate the temporal changes in treatment-seeking behavior from the 2014 and 2017/18 malaria indicator surveys (MIS). Additionally, the lack of a quality assured healthcare facility master list limits our ability to construct reporting rates which accurately account for facilities opening and closing during the study period. Ideally, this master list would also contain accurate information on the GPS coordinates for all healthcare facilities in the HMIS data, which would improve our geo-spatial model in Chapter 4.

Future directions for this project necessitate more data for the full analysis of the programmatic SMC campaign. Monthly routine case data from 2013, 2014, 2019, and 2020 would enable us to analyze the effectiveness of SMC on all rollout groups, meaning we could capture the impact of the full SMC campaign across the entire country. This would pose a novel contribution, as no study has examined the programmatic effectiveness of SMC on the 2014 or the 2019 rollout groups.

The fine-scale spatio-temporal model presented in Chapter 4 ameliorated some of the gaps in reporting quality from healthcare facilities, but limited our ability to analyze the malaria incidence since we had no data on the catchment populations of the health facilities. Data on the catchment populations or catchment areas would enable us to apply our model to estimate the effectiveness of SMC on reducing

the malaria incidence. Alternatively, data on the GPS coordinates for the remaining facilities not included in the Chapter 4 analysis would not only improve our model by incorporating more data, but would also enable us to utilize gravity models to estimate the catchment populations of health facilities manually.

Finally, our novel approach to modeling changes in the malaria burden through the malaria proportion of outpatient visits requires validation of the approach through its application in other settings. Differences-in-differences modeling with multiple treatment periods is an ongoing area of research, and our contribution can be extended to other disciplines whic utilize latent processes in differences-in-differences frameworks (Callaway & Sant'Anna, 2021; Huntington-Klein, 2021). Chapter 3 showed that the choice the malaria indicator analyzed to capture the effectiveness of SMC produces strong differences in the estimated effect size. Validation of our proposed indicator may be done through the analysis of other interventions in Burkina Faso using quasi-experimental approaches with HMIS routine case data. Alternatively, this methodology could be applied to routine case data in other high-burden countries to analyze the long-term trend in the malaria burden or to capture the effectiveness of interventions distributed through programmatic campaigns.

References

Adegboye, O., Gayawan, E., James, A., Adegboye, A., & Elfaki, F. (2021). Bayesian spatial modelling of ebola outbreaks in democratic republic of congo through the INLA-SPDE approach. *Zoonoses and Public Health*, *68*(5), 443–451.

Alegana, V. A., Okiro, E. A., & Snow, R. W. (2020). Routine data for malaria morbidity estimation in africa: Challenges and prospects. *BMC Medicine*, *18*, 1–13.

Amsalu, E., Liu, M., Li, Q., Wang, X., Tao, L., Liu, X., Luo, Y., Yang, X., Zhang, Y., Li, W.others. (2019). Spatial-temporal analysis of tuberculosis in the geriatric population of china: An analysis based on the bayesian conditional autoregressive model. *Archives of Gerontology and Geriatrics*, *83*, 328–337.

Ashton, R. A., Bennett, A., Al-Mafazy, A.-W., Abass, A. K., Msellem, M. I., McElroy, P., Kachur, S. P., Ali, A. S., Yukich, J., Eisele, T. P.others. (2019). Use of routine health information system data to evaluate impact of malaria control interventions in zanzibar, tanzania from 2000 to 2015. *EClinicalMedicine*, *12*, 11–19.

Ashton, R. A., Bennett, A., Yukich, J., Bhattarai, A., Keating, J., & Eisele, T. P. (2017). Methodological considerations for use of routine health information system data to evaluate malaria program impact in an era of declining malaria transmission. *The American Journal of Tropical Medicine and Hygiene*, *97*(3 Suppl), 46.

Baba, E., Hamade, P., Kivumbi, H., Marasciulo, M., Maxwell, K., Moroso, D., Roca-Feltrer, A., Sanogo, A., Johansson, J. S., Tibenderana, J.others. (2020). Effectiveness of seasonal malaria chemoprevention at scale in west and central africa: An observational study. *The Lancet*, *396*(10265), 1829–1840.

Baker, A. C., Larcker, D. F., & Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, *144*(2), 370–395.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2003). *Hierarchical modeling and analysis for spatial data*. Chapman; Hall/CRC.

Bertozzi-Villa, A., Bever, C. A., Koenker, H., Weiss, D. J., Vargas-Ruiz, C., Nandi, A. K., Gibson, H. S., Harris, J., Battle, K. E., Rumisha, S. F.others. (2021). Maps and metrics of insecticide-treated net access, use, and nets-per-capita in africa from 2000-2020. *Nature Communications*, *12*(1), 1–12.

Cairns, M., Ceesay, S. J., Sagara, I., Zongo, I., Kessely, H., Gamougam, K., Diallo, A., Ogboi, J. S., Moroso, D., Van Hulle, S.others. (2021). Effectiveness of seasonal malaria chemoprevention (SMC) treatments when SMC is implemented at scale: Case–control studies in 5 countries. *PLoS Medicine*, *18*(9), e1003727.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230.

Cameron, E., Young, A. J., Twohig, K. A., Pothin, E., Bhavnani, D., Dismer, A., Merilien, J. B., Hamre, K., Meyer, P., Le Menach, A.others. (2021). Mapping the endemicity and seasonality of clinical malaria for intervention targeting in haiti using routine case data. *Elife*, *10*, e62122.

CDC - malaria - about malaria - biology. (2020). In *Centers for Disease Control and Prevention*. Centers for Disease Control; Prevention. https://www.cdc.gov/malaria/about/biology/index.html

Chanda, E., Coleman, M., Kleinschmidt, I., Hemingway, J., Hamainza, B., Masaninga, F., Chanda-Kapata, P., Baboo, K. S., Dürrheim, D. N., & Coleman, M. (2012). Impact assessment of malaria vector control using routine surveillance data in zambia: Implications for monitoring and evaluation. *Malaria Journal*, *11*(1), 1–8.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition. *J. Off. Stat*, *6*(1), 3–73.

Dalrymple, U., Arambepola, R., Gething, P. W., & Cameron, E. (2018). How long do rapid diagnostic tests remain positive after anti-malarial treatment? *Malaria Journal*, *17*(1), 1–13.

Danwang, C., Khalil, É., Achu, D., Ateba, M., Abomabo, M., Souopgui, J., De Keukeleire, M., & Robert, A. (2021). Fine scale analysis of malaria incidence in under-5: Hierarchical bayesian spatio-temporal

modelling of routinely collected malaria data between 2012–2018 in cameroon. *Scientific Reports*, *11*(1), 1–10.

De Savigny, D., & Binka, F. (2004). Monitoring future impact on malaria burden in sub-saharan africa. *The American Journal of Tropical Medicine and Hygiene*, *71*(2 Supp), 224–231.

Diawara, F., Steinhardt, L. C., Mahamar, A., Traore, T., Kone, D. T., Diawara, H., Kamate, B., Kone, D., Diallo, M., Sadou, A.others. (2017). Measuring the impact of seasonal malaria chemoprevention as part of routine malaria control in kita, mali. *Malaria Journal*, *16*(1), 1–12.

Didan, K., Munoz, A. B., Solano, R., & Huete, A. (2015). MODIS vegetation index user's guide (MOD13 series). *University of Arizona: Vegetation Index and Phenology Lab*.

Druetz, T., Corneau-Tremblay, N., Millogo, T., Kouanda, S., Ly, A., Bicaba, A., & Haddad, S. (2018). Impact evaluation of seasonal malaria chemoprevention under routine program implementation: A quasi-experimental study in burkina faso. *The American Journal of Tropical Medicine and Hygiene*, *98*(2), 524.

Druetz, T., Ridde, V., Kouanda, S., Ly, A., Diabaté, S., & Haddad, S. (2015). Utilization of community health workers for malaria treatment: Results from a three-year panel study in the districts of kaya and zorgho, burkina faso. *Malaria Journal*, *14*(1), 1–12.

Ewing, V. L., Lalloo, D. G., Phiri, K. S., Roca-Feltrer, A., Mangham, L. J., & SanJoaquin, M. A. (2011). Seasonal and geographic differences in treatment-seeking and household cost of febrile illness among children in malawi. *Malaria Journal*, *10*(1), 1–8.

Fact sheet about malaria. (2022). In *World Health Organization*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/malaria

Galatas, B.others. (2021). *Strategic use of local information to target and implement malaria interventions sub-nationally in high endemic countries in africa*.

Gilbert, R. O. (1987). *Statistical methods for environmental pollution monitoring*. John Wiley & Sons.

Gómez-Rubio, V. (2020). *Bayesian inference with INLA*. CRC Press.

Hafen, R. (2016). Stlplus: Enhanced seasonal decomposition of time series by loess. *R Package Version 0.5*, *1*.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D.others. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049.

Huntington-Klein, N. (2021). *The effect: An introduction to research design and causality*. Chapman; Hall/CRC.

ICF International, National Center for Research and Training on Malaria (CNRFP) (Burkina Faso), National Institute of Statistics and Demography (Burkina Faso), National Program for the Fight Against Malaria (PNLP) (Burkina Faso). (2015). Burkina faso malaria indicator survey 2014. In *Malaria indicator survey, burkina faso*. Fairfax, United States of America: ICF International.

ICF International, National Center for Research and Training on Malaria (CNRFP) (Burkina Faso), National Institute of Statistics and Demography (Burkina Faso), National Program for the Fight Against Malaria (PNLP) (Burkina Faso). (2019). Burkina faso malaria indicator survey 2017-2018. In *Malaria indicator survey, burkina faso*. Fairfax, United States of America: ICF International.

Ingebrigtsen, R., Lindgren, F., & Steinsland, I. (2014). Spatial models with explanatory variables in the dependence structure. *Spatial Statistics*, *8*, 20–38.

Kirakoya-Samadoulougou, F., De Brouwere, V., Fokam, A. F., Ouédraogo, M., & Yé, Y. (2022). Assessing the effect of seasonal malaria chemoprevention on malaria burden among children under 5 years in burkina faso. *Malaria Journal*, *21*(1), 1–10.

Konaté, D., Diawara, S. I., Touré, M., Diakité, S. A., Guindo, A., Traoré, K., Diarra, A., Keita, B., Thiam, S., Keita, M.others. (2020). Effect of routine seasonal malaria chemoprevention on malaria trends in children under 5 years in dangassa, mali. *Malaria Journal*, *19*(1), 1–6.

Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., & Rue, H. (2018). *Advanced spatial modeling with stochastic partial differential equations using r and INLA*. Chapman; Hall/CRC.

Linard, C., Gilbert, M., Snow, R. W., Noor, A. M., & Tatem, A. J. (2012). Population distribution, settlement patterns and accessibility across africa in 2010. *PloS One*, *7*(2), e31743.

Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with r-INLA. *Journal of Statistical Software*, *63*, 1–25.

Malaria: The malaria vaccine implementation programme (MVIP). (2020). In *World Health Organization*. World Health Organization. https://www.who.int/news-room/q-a-detail/malaria-vaccine-implementation-programme

Malik, E. M., Hanafi, K., Ali, S. H., Ahmed, E. S., & Mohamed, K. A. (2006). Treatment-seeking behaviour for malaria in children under five years of age: Implication for home management in rural areas with high seasonal transmission in sudan. *Malaria Journal*, *5*(1), 1–5.

McCullagh, P., & Nelder, J. (1989). *Generalized linear models II*. Chapman; Hall, London.

Meremikwu, M. M., Donegan, S., Sinclair, D., Esu, E., & Oringanje, C. (2012). Intermittent preventive treatment for malaria in children living in areas with seasonal transmission. *Cochrane Database of Systematic Reviews*, *2*.

Ministere de la Sante du Burkina Faso. (2019). *Annuaire statistique 2018*.

Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time series missing value imputation in r. *R J.*, *9*(1), 207.

Müller, O., Traoré, C., Becher, H., & Kouyaté, B. (2003). Malaria morbidity, treatment-seeking behaviour, and mortality in a cohort of young children in rural burkina faso. *Tropical Medicine & International Health*, *8*(4), 290–296.

Newby, G., Hwang, J., Koita, K., Chen, I., Greenwood, B., Von Seidlein, L., Shanks, G. D., Slutsker, L., Kachur, S. P., Wegbreit, J.others. (2015). Review of mass drug administration for malaria and its operational challenges. *The American Journal of Tropical Medicine and Hygiene*, *93*(1), 125.

Nguyen, M., Howes, R. E., Lucas, T. C., Battle, K. E., Cameron, E., Gibson, H. S., Rozier, J., Keddie, S., Collins, E., Arambepola, R.others. (2020). Mapping malaria seasonality in madagascar using health facility data. *BMC Medicine*, *18*(1), 1–11.

Noor, A. M., Kinyoki, D. K., Mundia, C. W., Kabaria, C. W., Mutua, J. W., Alegana, V. A., Fall, I. S., & Snow, R. W. (2014). The changing risk of plasmodium falciparum malaria infection in africa: 2000–10: A spatial and temporal analysis of transmission intensity. *The Lancet*, *383*(9930), 1739–1747.

Ozodiegwu, I. D., Ambrose, M., Battle, K. E., Bever, C., Diallo, O., Galatas, B., Runge, M., & Gerardin, J. (2021). Beyond national indicators: Adapting the demographic and health surveys' sampling strategies and questions to better inform subnational malaria intervention policy. *Malaria Journal*, *20*(1), 1–7.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. C. (2007). Linear and nonlinear mixed effects models. *R Package Version*, *3*(57), 1–89.

Pohlert, T. (2016). Non-parametric trend tests and change-point detection. *CC BY-ND*, *4*.

Poirot, E., Skarbinski, J., Sinclair, D., Kachur, S. P., Slutsker, L., & Hwang, J. (2013). Mass drug administration for malaria. *Cochrane Database of Systematic Reviews*, *12*.

R Core Team and others. (2013). *R: A language and environment for statistical computing*.

Rasmussen, S. L. (2018). Plans and "off-plan activities": Exploring the roles of data and situated action in health planning in burkina faso. *The Electronic Journal of Information Systems in Developing Countries*, *84*(5), e12049.

Ridde, V., & Yaméogo, P. (2018). How burkina faso used evidence in deciding to launch its policy of free healthcare for children under five and women in 2016. *Palgrave Communications*, *4*(1), 1–9.

Rouamba, T., Samadoulougou, S., & Kirakoya-Samadoulougou, F. (2020). Addressing challenges in routine health data reporting in burkina faso through bayesian spatiotemporal prediction of weekly clinical malaria incidence. *Scientific Reports*, *10*(1), 1–15.

Rue, H., Martino, S., & Chopin, N. (2009). 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.

Scott, S., Cairns, M., Lal, S., Diallo, A., Ndiaye, J., Ouedraogo, J., Zongo, I., Doumagoum, D., Kessely, H., Bojang, K.others. (2018). Seasonal malaria chemoprevention coverage in seven west african countries, 2015–2016. *Report for UNITAID. London: London School of Hygiene & Tropical Medicine*.

Selvaraj, P., Wenger, E. A., & Gerardin, J. (2018). Seasonality and heterogeneity of malaria transmission determine success of interventions in high-endemic settings: A modeling study. *BMC Infectious Diseases*, *18*(1), 1–14.

Soma, D., Zogo, B., Taconet, P., Somé, A., Coulibaly, S., Baba-Moussa, L., Ouédraogo, G., Koffi, A., Pennetier, C., Dabiré, K.others. (2021). Quantifying and characterizing hourly human exposure to malaria vectors bites to address residual malaria transmission during dry and rainy seasons in rural southwest burkina faso. *BMC Public Health*, *21*(1), 1–9.

The PMI VectorLink Project, Abt Associates Inc. (2020). *PMI VectorLink burkina faso 2020 end of spray report: June 01 – 26, 2020*.

Tiono, A. B., Kangoye, D. T., Rehman, A. M., Kargougou, D. G., Kaboré, Y., Diarra, A., Ouedraogo, E., Nébié, I., Ouédraogo, A., Okech, B.others. (2014). Malaria incidence in children in south-west burkina faso: Comparison of active and passive case detection methods. *PloS One*, *9*(1), e86936.

Tipke, M., Louis, V. R., Yé, M., De Allegri, M., Beiersmann, C., Sié, A., Mueller, O., & Jahn, A. (2009). Access to malaria treatment in young children of rural burkina faso. *Malaria Journal*, *8*(1), 1–10.

Venables, W. N., & Ripley, B. D. (2013). *Modern applied statistics with s-PLUS*. Springer Science & Business Media.

Wagman, J., Cissé, I., Kone, D., Fomba, S., Eckert, E., Mihigo, J., Bankineza, E., Bah, M., Diallo, D., Gogue, C.others. (2020). Combining next-generation indoor residual spraying and drug-based malaria control strategies: Observational evidence of a combined effect in mali. *Malaria Journal*, *19*(1), 1–11.

Wangdi, K., & Clements, A. C. (2017). Spatial and temporal patterns of diarrhoea in bhutan 2003–2013. *BMC Infectious Diseases*, *17*(1), 1–9.

Wangdi, K., Clements, A. C., Du, T., & Nery, S. V. (2018). Spatial and temporal patterns of dengue infections in timor-leste, 2005–2013. *Parasites & Vectors*, *11*(1), 1–9.

Weiss, D. J., Lucas, T. C., Nguyen, M., Nandi, A. K., Bisanzio, D., Battle, K. E., Cameron, E., Twohig, K. A., Pfeffer, D. A., Rozier, J. A.others. (2019). Mapping the global prevalence, incidence, and mortality of plasmodium falciparum, 2000–17: A spatial and temporal modelling study. *The Lancet*, *394*(10195), 322–331.

Wilson, A. L., & Taskforce, I. (2011). A systematic review and meta-analysis of the efficacy and safety of intermittent preventive treatment of malaria in children (IPTc). *PloS One*, *6*(2), e16976.

World development indicators: Population ages 00-04 (% of total population). (2021). In *World Bank: DataBank*. https://databank.worldbank.org/source/world-development-indicators

World Health Organization. (2015). *Global technical strategy for malaria 2016-2030*. World Health Organization.

World Health Organization. (2020). *World malaria report 2020: 20 years of global progress and challenges*.

World Health Organization. (2021). *World malaria report 2021*.

World Health Organization and others. (2012). *WHO policy recommendation: Seasonal malaria chemoprevention (SMC) for plasmodium falciparum malaria control in highly seasonal transmission areas of the sahel sub-region in africa*. World Health Organization.

World Health Organization and others. (2018a). *High burden to high impact: A targeted malaria response*. World Health Organization.

World Health Organization and others. (2018b). *World health statistics 2018: Monitoring health for the SDGs, sustainable development goals*. World Health Organization.

Zombré, D., De Allegri, M., & Ridde, V. (2017). Immediate and sustained effects of user fee exemption on healthcare utilization among children under five in burkina faso: A controlled interrupted time-series analysis. *Social Science & Medicine*, *179*, 27–35.

APPENDIX A

Supplement to: Chapter 2

A.1. Data manipulation, adjustment factors, and analysis on individuals over 5 years old

A.1.1. Map of health districts in Burkina Faso

Figure A.1. Map of the 70 health districts in Burkina Faso with names. Colors correspond to the region to which each district belongs (Figure 2.1).

A.1.2. Flow chart for data procedure

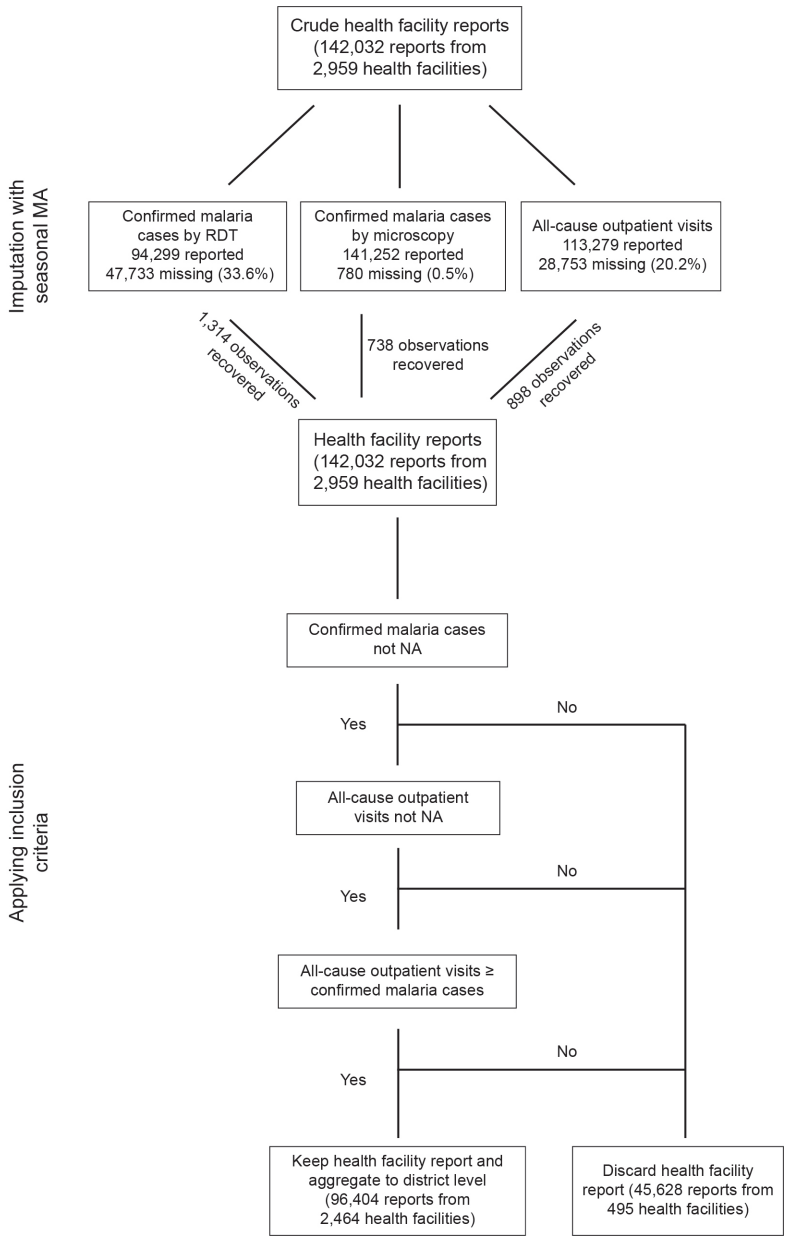


Figure A.2. Flow chart of data aggregation and imputation steps detailed in the methods section (see Chapter 2 methods). Between imputation and applying the inclusion criteria, health facility weights are calculated. After applying the inclusion criteria, health facility data is aggregated to the health district level. The same procedure was applied to individuals over 5 years old for the subsequent analyses on this age group.

A.1.3. Equation: Calculation of monthly weighted reporting rates per health district

Weighted reporting rates for each health district were calculated as follows. These formulas apply for monthly confirmed malaria cases and monthly all-cause outpatient visits, both of which will be referred to as "cases".

For each calendar month (m) (January through December), a health facility (h) has an average number of malaria cases ($\overline{C}_{h,m}$) reported for that month, across all years of data (y) (2015 to 2018) (Equation A.1.1).

$$\overline{C}_{h,m} = \frac{\sum_{y=2015}^{2018} C_{h,m,y}}{\sum_{y=2015}^{2018} \mathbf{I}[C_{h,m,y} \text{ not NA}]} \tag{A.1.1}$$

where $C_{h,m,y}$ is the number of cases for health facility h, calendar month m, and year y and $I[\cdot]$ is the indicator function equal to 1 if the event occurs, and 0 otherwise.

For each date (month m and year y pair), we modify the average number of cases $\overline{C}_{h,m}$ depending on whether the health facility is active/inactive (Equation A.1.2). The resulting average number of cases ($\overline{C}_{h,m,y}$) is either 0 if the health facility is inactive, and ($\overline{C}_{h,m}$) if the health facility is active during that date.

$$\overline{C}_{h,m,y} = \overline{C}_{h,m} \times I[h \text{ active for month } m \text{ in year } y] \tag{A.1.2}$$

The health facility weights ($w_{h,m,y}$) were calculated by dividing the average number of cases in health facility (h) reported for that month and year ($\overline{C}_{h,m,y}$) by the district sum of the average number of cases reported for that month and year for all health facilities (j) in a district ($\mathcal{D}$) (Equation A.1.3).

$$w_{h,m,y} = \frac{\overline{C}_{h,m,y}}{\sum_{j \in \mathcal{D}} \overline{C}_{j,m,y}} \tag{A.1.3}$$

The district monthly weighted reporting rate ($R_{m,y}$) was then calculated by summing the weights ($w_{h,m,y}$) of the active health facilities (h) that met the inclusion criteria each date (month m and year y) (Equation A.1.4).

(A.1.4) $$R_{m,y} = \sum h \in \mathcal{D} w_{h,m,y}$$

A.1.4. Equation: Equations to adjust malaria case counts

Equations to adjust monthly routine malaria cases for changes in health facility reporting rates and changes in treatment seeking rates.

First, crude malaria counts (C_0) are adjusted each date for changes in health facility reporting rates (R), either unweighted or weighted (see Chapter 2 methods and Equation A.1.5). This produces a malaria count adjusted for reporting rate (C_1) (Equation A.1.5).

(A.1.5) $$C_1 = \frac{C_0}{R}$$

Malaria counts are further adjusted for changes in treatment seeking rates (T), to produce a malaria count adjusted for reporting rate and treatment seeking rates (C_2) (Equation A.1.6).

(A.1.6) $$C_2 = \frac{C_1}{T}$$

A.1.5. Analysis of malaria incidence adjusted for unweighted reporting rate

Malaria incidence rates among children under 5 years old are adjusted for the crude, unweighted health facility reporting rate for each district. The time series for this adjusted malaria incidence rate (Figure A.3A) shows some difference to that of crude malaria incidence (Figure 2.2A), with malaria incidence showing a new spike the Bogodogo district in mid-2016. The overall trend shows an increasing trend (Figure A.3B), similar to that of crude malaria incidence and malaria incidence adjusted for weighted reporting rate (Figure 2.2C, 2.2D). The Sen's slope coefficient shows similar results to that of the crude malaria incidence (Figure 2.2E), with 41/70 districts showing an increasing trend, 1 district, Dori, with a decreasing trend, and 28/70 districts having no significant trend (Figure A.3C).

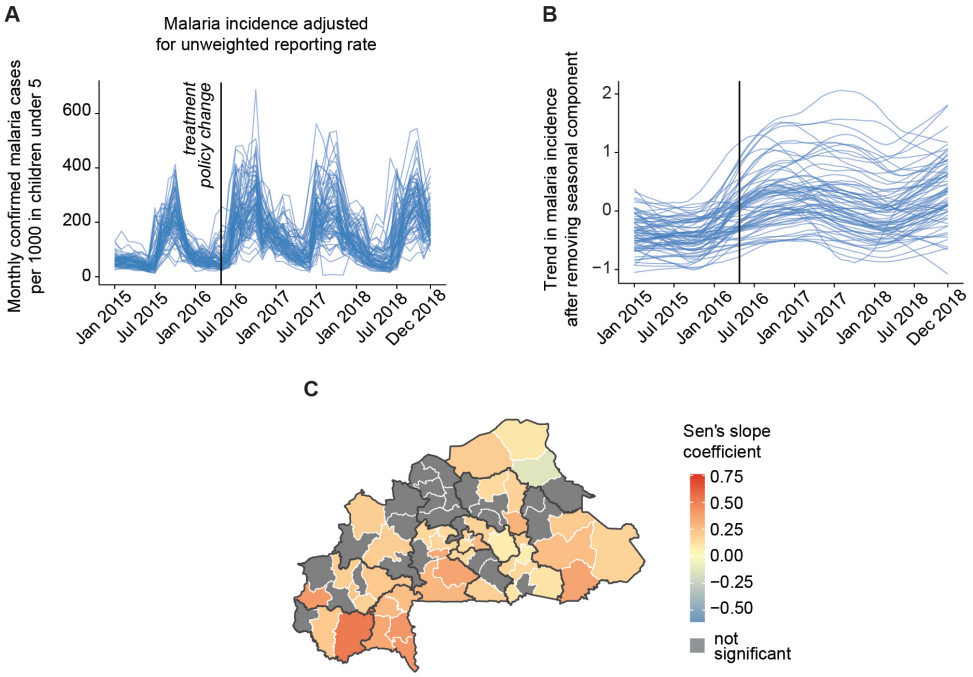


Figure A.3. Quantifying trends in the malaria incidence adjusted for unweighted reporting rate among children under 5 years old. A) Monthly confirmed malaria cases by RDT or microscopy per 1000 in children under 5 years old adjusted for each district's unweighted reporting rate. Each line represents one health district, and the black line indicates the 2016 policy change when treatment for childhood illnesses at public health facilities became free of charge. B) District-level trend components of the adjusted malaria incidence after seasonal trend decomposition. C) Sen's slope coefficient for the trend of the adjusted malaria incidence. Positive numbers indicate increasing trend. Gray: no significant increasing or decreasing trend according to the Mann-Kendall test.

A.1.6. Mapping Sen's slope coefficient for individuals over 5

We map the Sen's slope coefficient for the time series for the 5 indicator variables considered among individuals over 5 years old across the 70 health districts. Crude malaria incidence shows 34 districts with an increasing trend, 2 districts with a decreasing trend, and 34 districts with a non-significant direction of the trend component according to the Mann-Kendall test. Malaria adjusted for weighted reporting rate shows 31 districts with an increasing trend, 2 districts with a decreasing trend, and 37 districts with a non-significant direction of the trend component. All-cause outpatient visits showed an increasing trend in 58 districts, a decreasing trend in no districts, and a non-significant direction of the trend component in 12 districts. Non-malarial outpatient visits showed an increasing trend in 60 districts, a decreasing trend in 2 districts, and a non-significant direction of the trend component in 8 districts. Finally, the malaria proportion of outpatient visits showed an increasing trend in 3 districts, a decreasing trend in 25 districts, and a non-significant direction of the trend component in 42 districts.

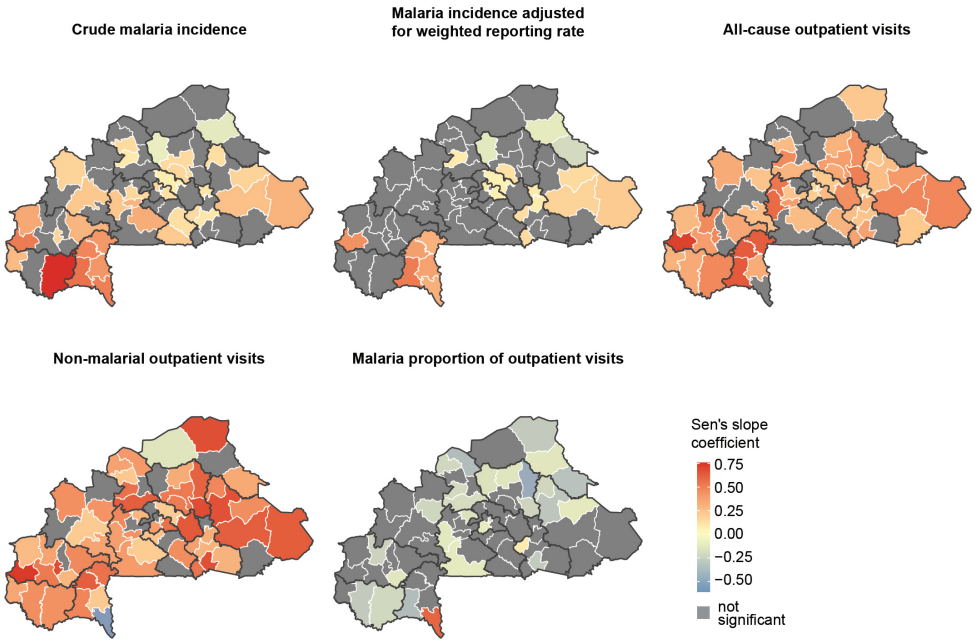


Figure A.4. Mapping the Sen's slope coefficient for the trends in the 5 indicator variables considered among individuals over 5 years old. Positive numbers indicate increasing trend. Gray: no significant increasing or decreasing trend according to the Mann-Kendall test.

A.1.7. Malaria testing among children under 5

Tested malaria cases among children under 5 years old come from RDT tests (> 99.5 of tests) and microscopy tests (< 0.5 of tests). The time series of tested cases shows a slight upward trend with seasonal peaks during each rainy season. The high malaria transmission season in 2016 sees the highest number of tested cases (Figure A.5A). The extracted trend components show the increasing trend, with the sharpest increase occurring between 2015 and the high transmission season of 2016. There is a slight reduction in the trend in 2017 after the spike in testing in 2016 (Figure A.5B). This increase in testing could correspond to the change in treatment policy for children under 5, leading to increased treatment-seeking for febrile illness. The Sen's slope coefficient shows that 51/70 districts have an increasing trend in the testing rate, with 19/70 districts having no significant trend (Figure A.5C).

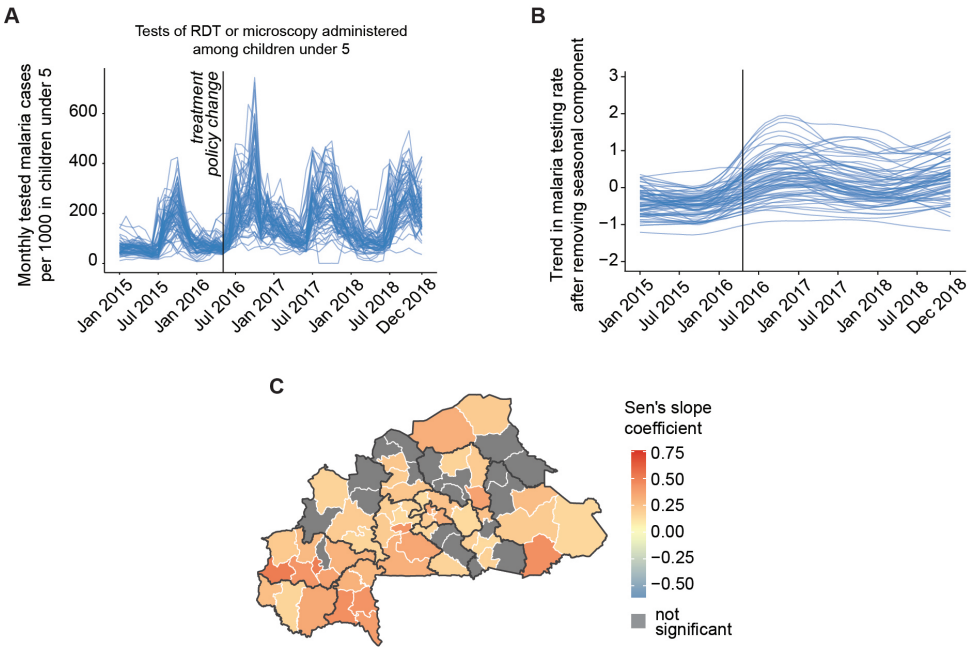


Figure A.5. Quantifying trends in the malaria testing rate among children under 5 years old. A) Monthly tested malaria cases by RDT or microscopy per 1000 in children under 5 years old. Each line represents one health district, and the black line indicates the 2016 policy change when treatment for childhood illnesses at public health facilities became free of charge. B) District-level trend components of the malaria testing rate after seasonal trend decomposition. C) Sen's slope coefficient for the trend of the malaria testing rate. Positive numbers indicate increasing trend. Gray: no significant increasing or decreasing trend according to the Mann-Kendall test.

Testing and confirmation numbers from RDT and microscopy tests are used in the analysis. The time series of the test positivity rate (TPR) are seasonal with low values occurring during the dry season each year and peaks of about 90% during the wet, high transmission season for malaria each year (Figure A.6A). The trend component shows an overall decrease over the time period. The timing of the 2016 change in healthcare policy did not seem to affect the trend of this variable meaningfully (Figure A.6B). Mapping the Sen's slope coefficient shows that 50/70 districts have a decreasing trend, 19 districts have a non-significant direction of the trend component, and one district, Mangodara in the Cascades region, has an increasing trend in the TPR (Figure A.6C).

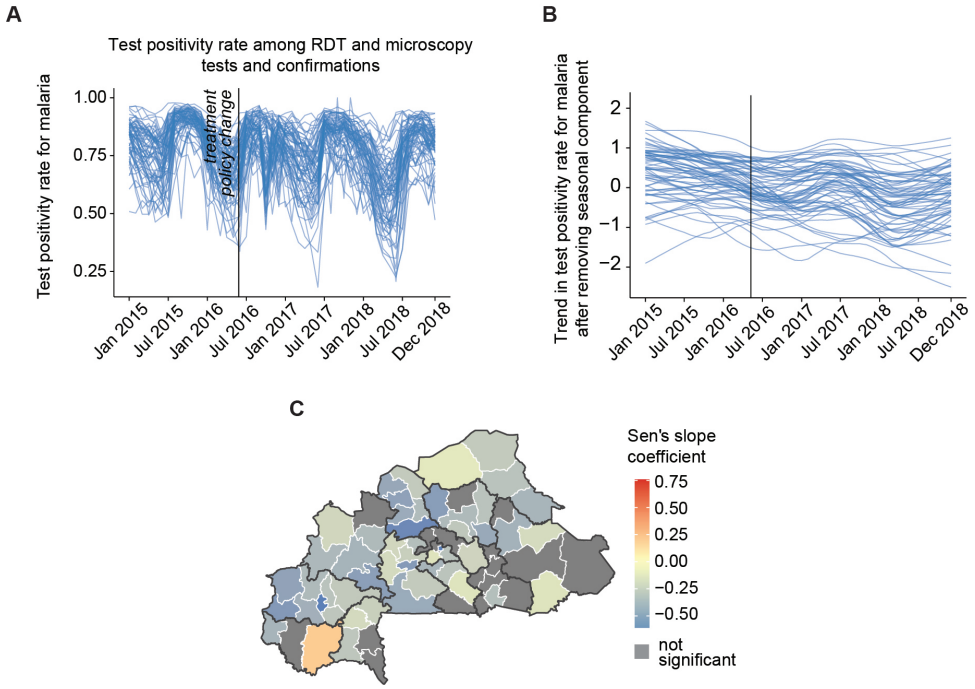
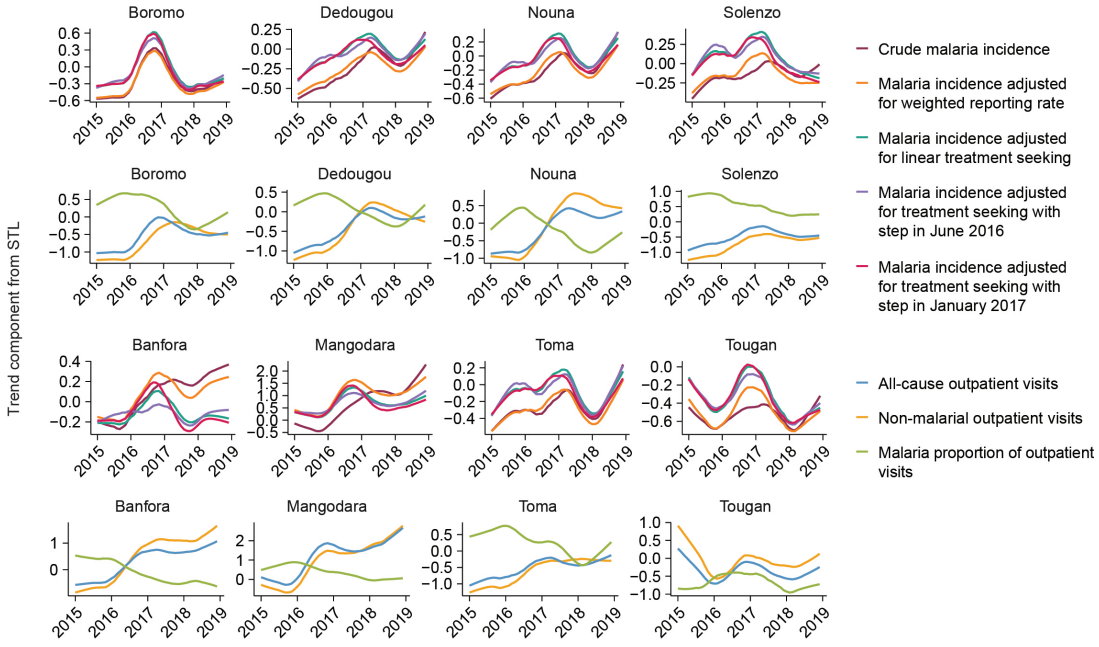
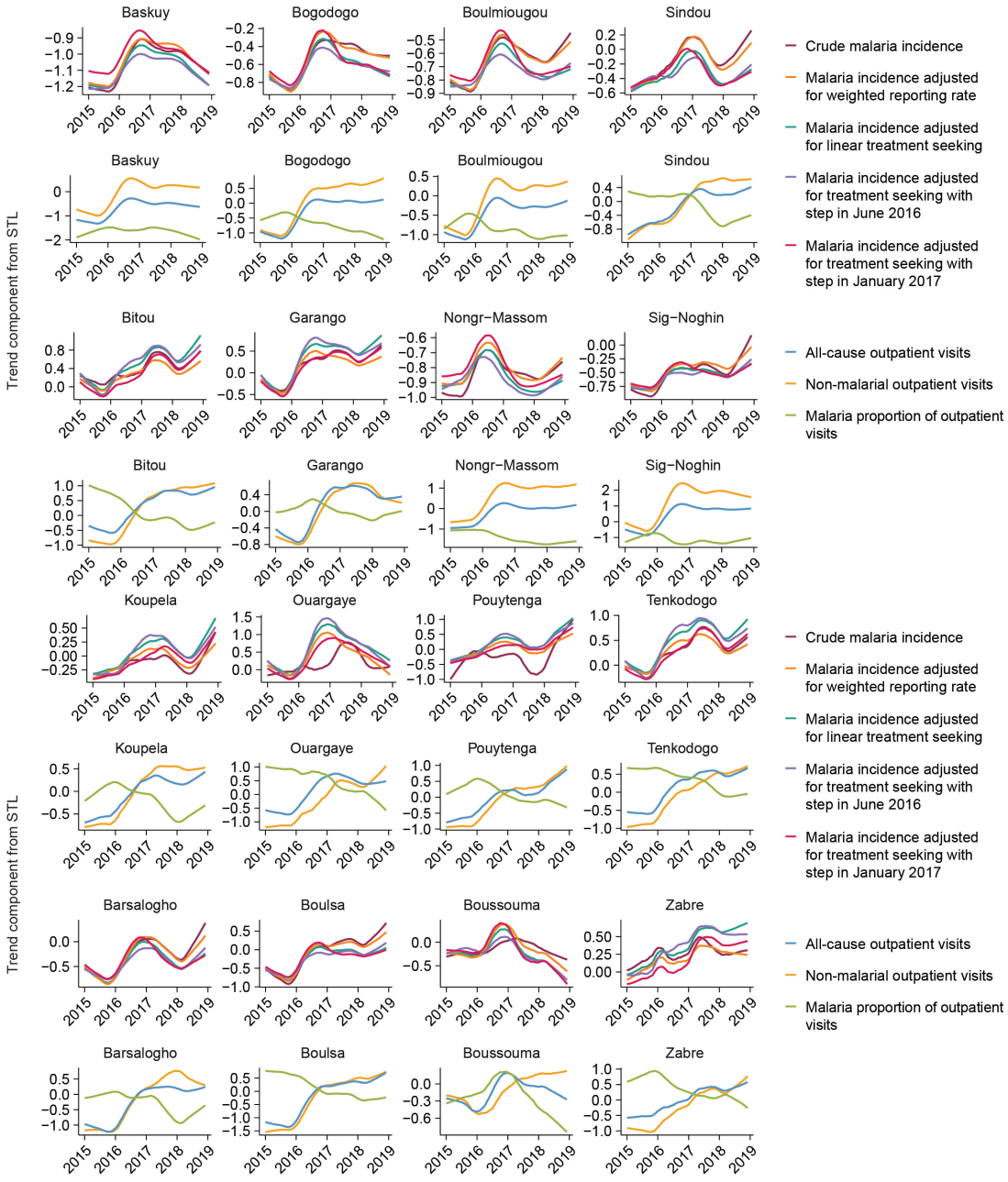


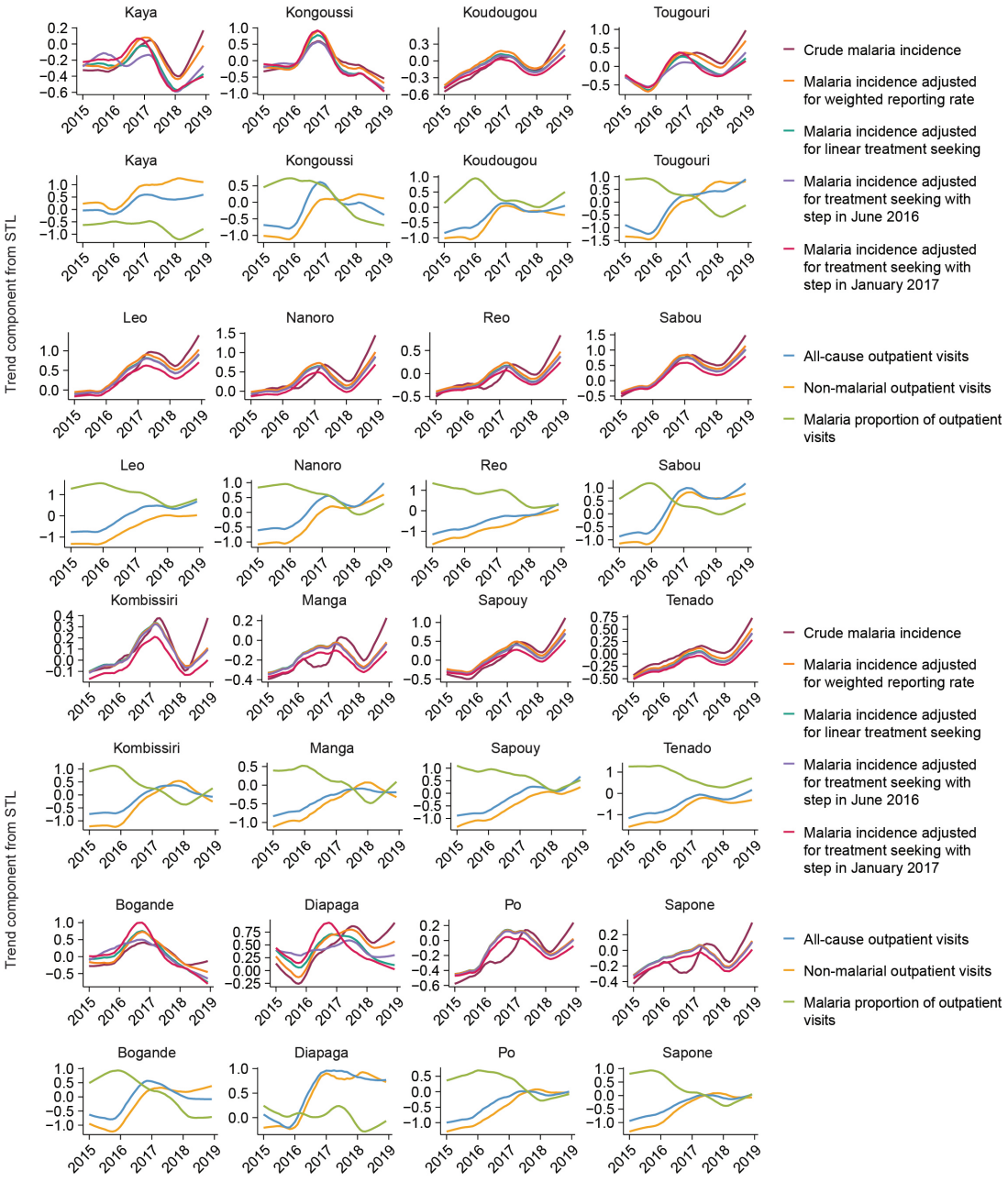
Figure A.6. Quantifying trends in the test positivity rate for malaria among children under 5 years old. Tests from both RDT and microscopy techniques are considered in this analysis. A) Monthly test positivity rate in children under 5 years old. Each line represents one health district, and the black line indicates the 2016 policy change when treatment for childhood illnesses at public health facilities became free of charge. B) District-level trend components of the test positivity rate after seasonal trend decomposition. C) Sen's slope coefficient for the trend of the test positivity rate. Positive numbers indicate increasing trend. Gray: no significant increasing or decreasing trend according to the Mann-Kendall test.

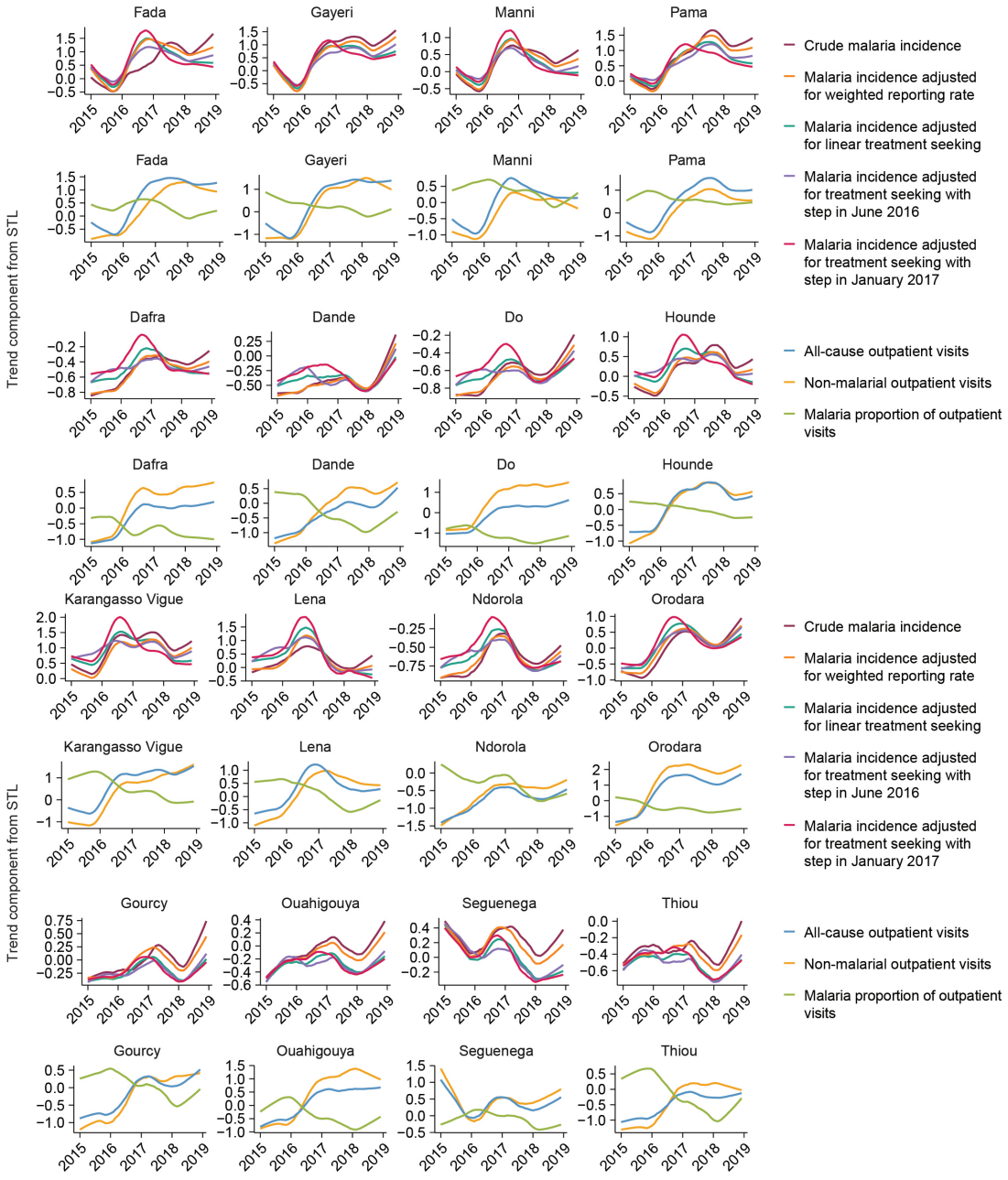
A.1.8. Analysis of trends in all 8 outcome variables for children under 5 across all 70 health districts in Burkina Faso

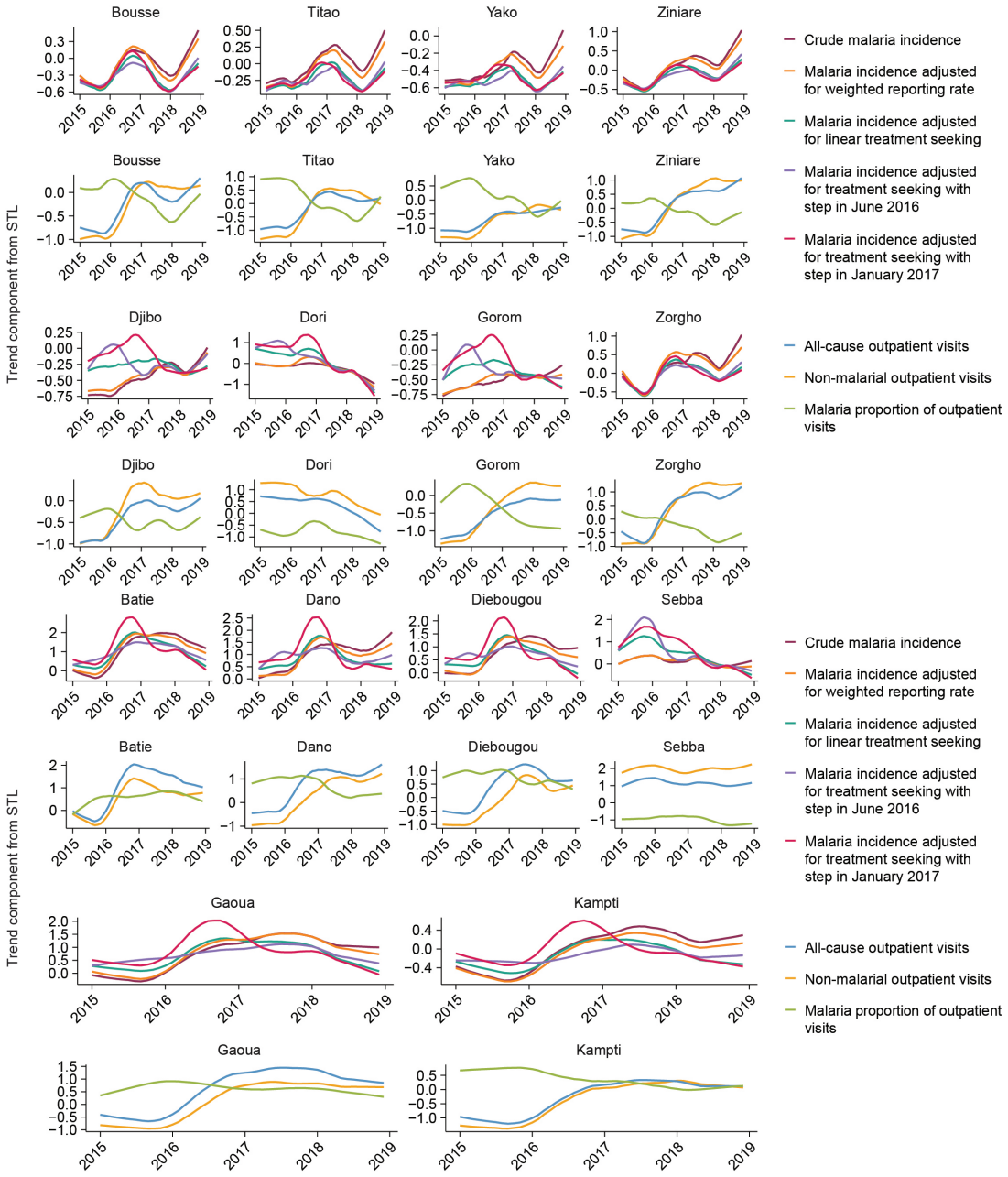
The differences in the trend components from the seasonal trend decomposition across the 8 outcome variables (crude malaria incidence, malaria incidence adjusted for weighted reporting rate, malaria incidence adjusted for linear treatment seeking, malaria incidence adjusted for treatment seeking with a step in June 2016, malaria incidence adjusted for treatment seeking with a step in January 2017, all-cause outpatient visits, non-malarial outpatient visits, and the malaria proportion of outpatient visits) is heterogeneous across the 70 districts.




Baskuy
Bogodogo
Boulmiougou
Sindou
Bitou
Garango
Nongr−Massom
Sig−Noghin
Koupela
Ouargaye
Pouytenga
Tenkodogo
Barsalogho
Boulsa
Boussouma
Zabre
Trend component from STL
2015
2016
2017
2018
2019
Crude malaria incidence
Malaria incidence adjusted for weighted reporting rate
Malaria incidence adjusted for linear treatment seeking
Malaria incidence adjusted for treatment seeking with step in June 2016
Malaria incidence adjusted for treatment seeking with step in January 2017
All-cause outpatient visits
Non-malarial outpatient visits
Malaria proportion of outpatient visits


Kaya
Kongoussi
Koudougou
Tougouri
Leo
Nanoro
Reo
Sabou
Kombissiri
Manga
Sapouy
Tenado
Bogande
Diapaga
Po
Sapone
Trend component from STL
Crude malaria incidence
Malaria incidence adjusted for weighted reporting rate
Malaria incidence adjusted for linear treatment seeking
Malaria incidence adjusted for treatment seeking with step in June 2016
Malaria incidence adjusted for treatment seeking with step in January 2017
All-cause outpatient visits
Non-malarial outpatient visits
Malaria proportion of outpatient visits


Fada
Gayeri
Manni
Pama
Dafra
Dande
Do
Hounde
Karangasso Vigue
Lena
Ndorola
Orodara
Gourcy
Ouahigouya
Seguenega
Thiou
Trend component from STL
Crude malaria incidence
Malaria incidence adjusted for weighted reporting rate
Malaria incidence adjusted for linear treatment seeking
Malaria incidence adjusted for treatment seeking with step in June 2016
Malaria incidence adjusted for treatment seeking with step in January 2017
All-cause outpatient visits
Non-malarial outpatient visits
Malaria proportion of outpatient visits


Bousse
Titao
Yako
Ziniare
Djibo
Dori
Gorom
Zorgho
Batie
Dano
Diebougou
Sebba
Gaoua
Kampti
Trend component from STL
Crude malaria incidence
Malaria incidence adjusted for weighted reporting rate
Malaria incidence adjusted for linear treatment seeking
Malaria incidence adjusted for treatment seeking with step in June 2016
Malaria incidence adjusted for treatment seeking with step in January 2017
All-cause outpatient visits
Non-malarial outpatient visits
Malaria proportion of outpatient visits

A.2. Spatial model for small-area estimation on febrile treatment-seeking rates

A.2.1. Introduction

We create a small-area estimation model to capture the fine-scale rate of treatment seeking to health facilities for children under 5 years old in Burkina Faso from Malaria Indicator Survey (MIS) data. The MIS defines treatment seeking as seeking treatment for a fever within the past 2 weeks to any public or private health facility (ICF International, 2015, 2019). The MIS is powered at the Regional level, thus to estimate the treatment-seeking rate at a finer spatial resolution (at the pixel-level), we use the Integrated Nested Laplacian Approximation (INLA) method in a generalized linear mixed-effects modeling framework (Ingebrigtsen et al., 2014; Lindgren & Rue, 2015; Rue et al., 2009).

We use cluster-level data from the 2014 and 2017/18 MIS to understand the rate of treatment seeking at the pixel-level. We create a mesh so that each pixel represents a 1×1 km square. The long gap between the surveys and the change of healthcare policy in 2016, which made treatment for childhood illnesses free, prohibits the use of data from both surveys in a spatio-temporal model.

Children in clusters are interviewed in different months, even within the same cluster. For the 2014 MIS, children are interviewed in September, October, November, and December. For the 2017/18 MIS, children are interviewed in November and December of 2017, and January, February, and March of 2018.

If a cluster has children within it interviewed in different months, we take the month with the maximum number of children interviewed to be the "interview date" for that cluster. With this method, all clusters have "interview dates" with at least 50% of the children interviewed within the selected month. For 2014, only 6/248 clusters have "interview dates" with less than 80% of children interviewed, and 16/248 with less than 90% of children interviewed. For 2017/18, only 3/245 clusters have "interview dates" with less than 90%, and 5/245 clusters with less than 100% of children interviewed.

A.2.2. Model

We create a separate model for estimating treatment seeking in 2014 and in 2017/18 from both MIS surveys. Our model is a spatial generalized mixed-effects model, with structured and unstructured spatial random effects. We apply the INLA methodology to fit the model using the **R-INLA** package (Krainski et al., 2018). We use a stochastic partial differential equations (SPDE) specification with a Matern covariance function for the structured spatial random effect, and a district-level random effect with a general i.i.d. Gaussian distribution for the unstructured random effect (Lindgren & Rue, 2015; Rue et al., 2009).

A.2.2.1. Covariates. We explore the inclusion of climate fixed effects covariates in monthly precipitation (Hersbach et al., 2020), monthly surface air temperature (Hersbach et al., 2020), and NDVI (Didan et al., 2015). We include precipitation and air temperature at 4 different time lags: lag 0, lag 1, lag 2, and lag 3. We also explore the use of a population surface (Linard et al., 2012) and a modeled surface of walking travel time to public health facilities (Weiss et al., 2019) as additional fixed effects. We take the log-transform of all variables considered. These covariates and their sources are summarized in table A.1.

Covariate	**Description**	**Source**
Precipitation	Monthly average precipitation	ERA5 (ERA5, 2022)
Air temperature	Monthly average surface air temperature	ERA5 (ERA5, 2022)
Travel time surface	Estimated walking travel time to nearest health facility	The Malaria Atlas Project (Weiss et al., 2019)
NDVI	Monthly normalized difference vegetation index	NASA EarthData (EarthData, 2022)
Population	Pixel-level number of people per cell (1 km $\times$ 1 km)	WorldPop (WorldPop, 2022)

Table A.1. Table of covariates considered as fixed effects. These log-transformed covariates are considered in both the 2014 and 2017/18 models.

We include covariates in a forward step-wise selection method to pick the significant fixed effects to include in the final model. We explore grouping similar measurements to reduce variation in the climate covariates between clusters. When a climate covariate is significant, we group the covariates using the

"quantiles" method of the inla.group function which is part of the R-INLA package (Krainski et al., 2018). We use a grid-search approach to select the optimal group size (between $2-20$ groups) in a model with only the grouped predictor; the optimal size is chosen according to the lowest value of the negative sum of the log CPO (SLCPO).

The final set of covariates used for the 2014 model and the 2017/18 model are shown in table A.2 below.

Model for 2014 data		**Model for 2017/18 data**	
Covariate	Grouping	Covariate	Grouping
Travel time surface	None	Precipitation (lag 1)	7 groups
Air temperature (lag 1)	4 groups		

Table A.2. Table of final covariates considered as fixed effects for each model. Grouping refers to the final number of groups created for the covariate using the **inla.group** function.

A.2.2.2. Model formulation. Let $Y_{j\mathcal{D}}$ be the number of children who seek medical treatment for a recent fever in cluster j which is located inside health district $\mathcal{D}$. $N_{j\mathcal{D}}$ is the total number of children who had a recent fever in that cluster. We assume that $Y_{j\mathcal{D}}$ follows a Binomial distribution, $Y_{j\mathcal{D}}|N_{j\mathcal{D}} \sim \text{Bin}(N_{j\mathcal{D}}, \pi_{j\mathcal{D}})$, where $\pi_{j\mathcal{D}}$ is the proportion of children who seek medical treatment for a recent fever.

We model the latent process $\eta_{j\mathcal{D}} = \text{logit}(\pi_{j\mathcal{D}})$ with an SPDE with the following specification

$$\begin{aligned} Y_{j\mathcal{D}} N_{j\mathcal{D}} &\sim \text{Bin}(N_{j\mathcal{D}}, \pi_{j\mathcal{D}}) \\ \eta_{j\mathcal{D}} &= X_j\beta + v_{\mathcal{D}} + u_j \\ \mathbf{v}|\tau &\sim \text{N}(0, \Sigma_{\mathbf{v}}) \\ \mathbf{u}|\theta &\sim \text{GF}(0, \Sigma_{\mathbf{u}}) \end{aligned} \tag{A.2.1}$$

where X_j is the matrix of covariates (shown in table A.2) for the model at cluster location j, $v_{\mathcal{D}}$ is the random effect associated with the district $\mathcal{D}$ to which cluster j belongs with an i.i.d. Gaussian covariance structure, and u_j are structured spatial random effects with a Matern covariance structure.

The district random effect $\mathbf{v} = (v_1, \ldots, v_n)$ has a Gaussian distribution with a general i.i.d. covariance structure with hyperparameter τ. The SPDE random effect $\mathbf{u} = (u_1, \ldots, u_m)$ has Gaussian random field process that captures the fine-scale spatial dependence. For the spatial correlation structure $\Sigma_{\mathbf{u}}$, we use a Matern covariance function, which enables the SPDE (Krainski et al., 2018). This assumes that the correlation between two clusters j and j' decreases as the distance between the two clusters increases, and vice versa (Ingebrigtsen et al., 2014; Lindgren & Rue, 2015).

The models are evaluated according to the negative sum of the log CPO (SLCPO) and the deviance information criterion (DIC). Additionally, we measure the mean absolute error (MAE) to the regional estimates of the treatment-seeking rate reported in the MIS. We aggregate results to the region level using the population weighted average of the pixel-level values contained in each region. The population surface used in this aggregation is that from WorldPop (Linard et al., 2012).

District-level estimates are estimated by using the population weighted average of the pixel-level values contained in each health district, where the population surface used is from World-Pop.

A.2.3. Results

A.2.3.1. Model selection & intermediate results. We present the results for the intermediate models which were used to create the final model. We evaluate the models based on the significance of their log-transformed covariates, the model SLCPO, and DIC values.

The intermediate results for the 2014 survey year show that travel time to health facilities, precipitation at lag 1 and lag 2, and air temperature at lag 1 and lag 2 are significant in the models fitted with only one covariate (Table A.3). Including both lagged terms for precipitation and air temperature, respectively, shows only significance in the lag 1 variable for both climate covariates. Looking at the model with travel time, precipitation at lag 1, and air temperature at lag 1, we see that lagged precipitation is no longer significant in this model. We arrive at our final model, which includes travel time and lagged air temperature, by reducing the variability in the climate covariate by grouping the lagged air temperature (Table A.5), which has a smaller DIC value and a smaller MAE to the regional estimates from the MIS survey.

The intermediate results for the 2017/18 survey year show that only precipitation at lag 1 is significant in the models fitted with only one covariate (Table A.4). The travel time to health facilities is no longer significant for this survey year. The final model contains precipitation at lag 1 after reducing the variability in the covariate by grouping (Table A.5).

A.2.3.2. Model summaries. We show the final results for the 2014 and 2017/18 models below. Table A.5 shows the model summary statistics for the fixed effects and the model diagnostic metrics. The model for the 2014 data contains the walking travel time and the air temperature of the previous month (lag 1) grouped into 4 segments as fixed effects. The range of the walking travel time to health facilities is from $0 - 6.7$ hours, with the average being roughly 1 hour and roughly 75% of clusters being within 1.34 hours of the nearest health facility. The monthly air temperature for the dates analyzed ranges from $25.4° - 31.7°$ C, with the average temperature being 27.8°C. The negative coefficient on the travel time coefficient shows that longer walking times to health facilities decrease the treatment-seeking rate. Similarly, the negative coefficient on air temperature indicates that higher temperatures in the month previous to the interview diminish the treatment-seeking rates of children. The model for the 2017/18 data only contains the average precipitation of the previous month (lag 1) grouped into 7 segments as the fixed effect. The range of the precipitation values for the dates analyzed is from $0 - 21.9$ cm. The negative coefficient on the precipitation coefficient in this model means that higher values of precipitation in the month previous to the interview lead to lower values of treatment-seeking rate for children.

We show the modeled surface with pixel-level predicted values of the treatment-seeking rate along with the pixel-level standard deviation from the posterior distribution. We also show the district-level and regional-level aggregated fitted values. These results are aggregated using a population weighted aggregation methodology, where the pixel-level population is derived from the modeled population surface provided by WorldPop.

A.2.3.3. Results for 2014. The results for the final model for the 2014 survey year are shown in Figure A.7 with the pixel-level estimated values. We see regions of low treatment seeking in more rural areas such as the Nord region in the north and the south western regions of Hauts Bassins and Sud-Ouest (Figure A.7A). Looking at the uncertainty surface, we see high standard deviation values for the posterior

Fixed effects	**Summary fixed effects (mean [0.025, 0.975])**	**-SLCPO**	**DIC**	**MAE**
β_0	0.145 [-0.169, 0.459]	492.47	951.64	0.033
β_0 precip.	0.163 [-0.144, 0.470] 0.117 [-0.036, 0.269]	492.78	952.68	0.029
β_0 precip. (lag1)	0.194 [-0.088, 0.475] 0.203 [0.054, 0.353]	487.98	950.08	0.035
β_0 precip. (lag2)	0.198 [-0.081, 0.477] 0.229 [0.072, 0.388]	491.78	948.98	0.038
β_0 precip. (lag3)	0.14 [-0.179, 0.458] 0.07 [-0.082, 0.222]	493.30	951.19	0.038
β_0 precip. (lag1) precip. (lag2)	0.202 [-0.078, 0.482] 0.088 [-0.139, 0.314] 0.16 [-0.075, 0.398]	493.05	950.19	0.038
β_0 air temp.	0.077 [-0.311, 0.465] 0.159 [-0.003, 0.323]	491.16	947.77	0.041
β_0 air temp. (lag1)	0.205 [-0.065, 0.475] -0.251 [-0.402, -0.100]	489.25	949.04	0.036
β_0 air temp. (lag2)	0.191 [-0.119, 0.501] -0.29 [-0.502, -0.079]	490.23	947.68	0.037
β_0 air temp. (lag3)	0.16 [-0.141, 0.461] -0.119 [-0.293, 0.055]	493.08	950.71	0.039
β_0 air temp. (lag1) air temp. (lag2)	0.205 [-0.105, 0.515] -0.191 [-0.365, -0.018] -0.173 [-0.409, 0.06]	488.86	947.47	0.036
β_0 travel time	0.253 [-0.045, 0.551] -0.27 [-0.388, -0.151]	482.92	939.92	0.052
β_0 NDVI	0.144 [-0.159, 0.447] 0.123 [-0.039, 0.285]	492.36	952.1	0.032
β_0 travel time precip. (lag1) air temp. (lag1)	0.291 [-0.01, 0.591] -0.276 [-0.391, -0.161] -0.347 [-0.798, 0.105] -0.594 [-1.058, -0.129]	478.24	934.82	0.032
β_0 travel time air temp. (lag1)	0.296 [0.004, 0.587] -0.27 [-0.386, -0.155] -0.254 [-0.401, -0.108]	478.42	935.18	0.030

Table A.3. Table of intermediate models for the 2014 survey year. These models were discarded based on the significance of their log-transformed covariates, the model SLCPO, and DIC values.

Fixed effects	**Summary fixed effects (mean [0.025, 0.975])**	**-SLCPO**	**DIC**	**MAE**
β_0	1.188 [0.905, 1.472]	281.93	523.43	0.041
β_0 precip.	1.19 [0.9, 1.48] -0.01 [-0.249, 0.229]	282.85	524.36	0.041
β_0 precip. (lag1)	1.183 [0.88, 1.487] -0.172 [-0.379, -0.04]	280.96	521.55	0.033
β_0 precip. (lag2)	1.196 [0.908, 1.485] 0.013 [-0.255, 0.295]	283.52	524.93	0.042
β_0 precip. (lag3)	1.236 [0.939, 1.534] 0.176 [-0.143, 0.501]	282.52	523.77	0.049
β_0 air temp.	1.213 [0.927, 1.5] -0.189 [-0.423, 0.041]	281.28	522.44	0.046
β_0 air temp. (lag1)	1.161 [0.867, 1.456] -0.215 [-0.468, 0.039]	281.55	521.47	0.036
β_0 air temp. (lag2)	1.206 [0.905, 1.508] 0.124 [-0.134, 0.383]	282.08	523.4	0.045
β_0 air temp. (lag3)	1.177 [0.863, 1.492] 0.183 [-0.033, 0.401]	280.81	521.02	0.037
β_0 NDVI	1.186 [0.893, 1.48] 0.056 [-0.216, 0.331]	283.09	523.9118	0.038
β_0 travel time	1.219 [0.937, 1.503] -0.12 [-0.328, 0.088]	283.13	524.34	0.045

Table A.4. Table of intermediate models for the 2017/18 survey year. These models were discarded based on the significance of their log-transformed covariates, the model SLCPO, and DIC values.

predicted treatment seeking estimates in the borders of the south western regions and in the Est region in the east, where there are few clusters samples as part of the MIS (Figure A.7B). The lowest uncertainty is seen in the center regions, where the majority of sampled clusters are located.

The district-level aggregated estimates using a population weighted average are shown in Figure A.8A. We see that there is more variation across neighboring districts when using the crude MIS estimates (Figure A.8B) which are not powered to produce district-level estimates. Our estimates produce a

	Fixed effects	Summary fixed effects (mean [0.025, 0.975])	-SLCPO	DIC	MAE
Model for 2014 data	β_0	0.285 [0.028, 0.541]	477.24	934.52	0.030
	travel time	-0.265 [-0.379, -0.152]			
	grouped air temp. (lag 1)	-0.278 [-0.425, -0.131]			
Model for 2017/18 data	β_0	1.154 [0.852, 1.457]	280.17	520.57	0.033
	grouped precip. (lag 1)	-0.302 [-0.598, -0.006]			

Table A.5. Model results for the 2014 and 2017/18 survey years. Fixed effects are summarized by the mean posterior value along with the 95% confidence interval. Model diagnostic metrics are shown with the negative log of the sum CPO (SLCPO), the DIC, and the MAE to the regional estimates from the MIS.

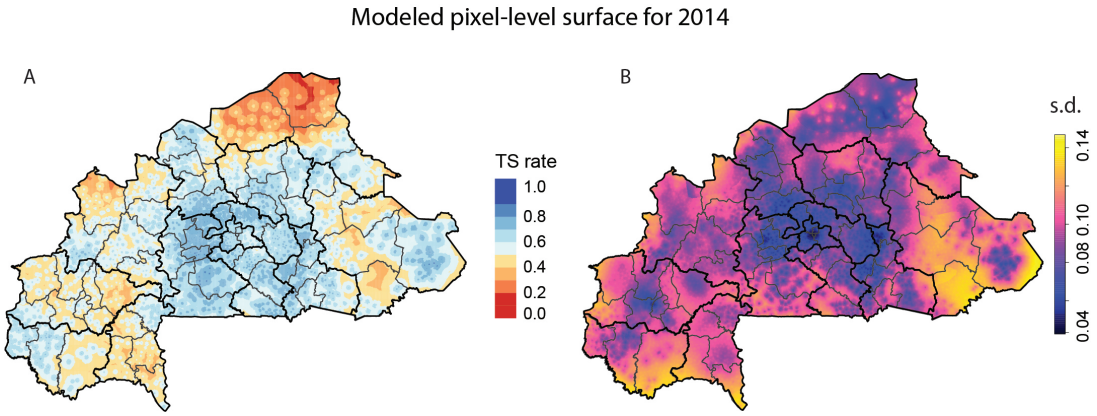


Figure A.7. Pixel-level surface for 2014 modeled using SPDE-INLA model. A) Pixel-level treatment-seeking (TS) rate estimated from the model where the value shown is the mean of the posterior predicted value from the model. B) Uncertainty surface from the model where the value shown is the standard deviation (s.d.) of the posterior predicted values.

smoother view of treatment-seeking behavior across Burkina Faso, and fill in the gaps in districts which contain no sampled clusters (shown in gray in Figure A.8B).

A.2.3.4. Results for 2017/18. Results for the final model for the 2017/18 survey year are shown in Figure 3 with the pixel-level estimated values. The estimated treatment-seeking rates are higher than from the 2014 survey year, as expected, with a few locations of lower treatment seeking (Figure A.9A). We see high model uncertainty in areas outside the center regions, and areas in the south western regions. The Nord, Est, and Boucle du Mouhoun regions have some of the highest high standard deviation values. The higher model uncertainty for the 2017/18 survey year than for the 2014 survey year may be driven

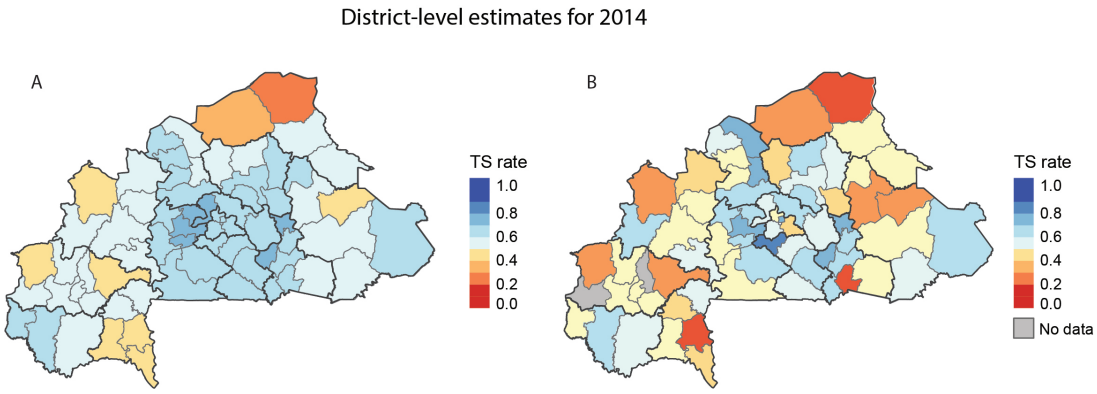


Figure A.8. District-level aggregated surface for 2014. A) District-level treatment-seeking (TS) rate aggregated using the pixel-level estimates from the model. B) District-level estimates from the 2014 MIS. Survey is not powered at the district level, so some districts have no estimated data.

by the dryer season in which the 2017/18 survey was conducted, which has lower malaria transmission rates than the rainy season in which the 2014 survey was conducted.

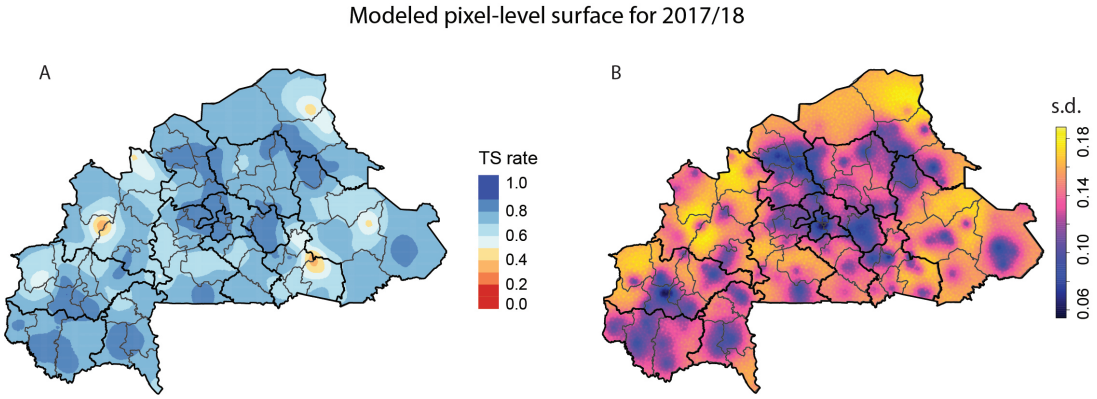


Figure A.9. Pixel-level surface for 2017/18 modeled using SPDE-INLA model. A) Pixel-level treatment-seeking (TS) rate estimated from the model where the value shown is the mean of the posterior predicted value from the model. B) Uncertainty surface from the model where the value shown is the standard deviation (s.d.) of the posterior predicted values.

The district-level aggregated estimates are shown in Figure A.10A, and show high treatment seeking rates across the country. The large difference seen between the modeled values and the estimated values from the 2017/18 MIS (Figure A.10B may be explained by a low number of clusters sampled in some of those districts.

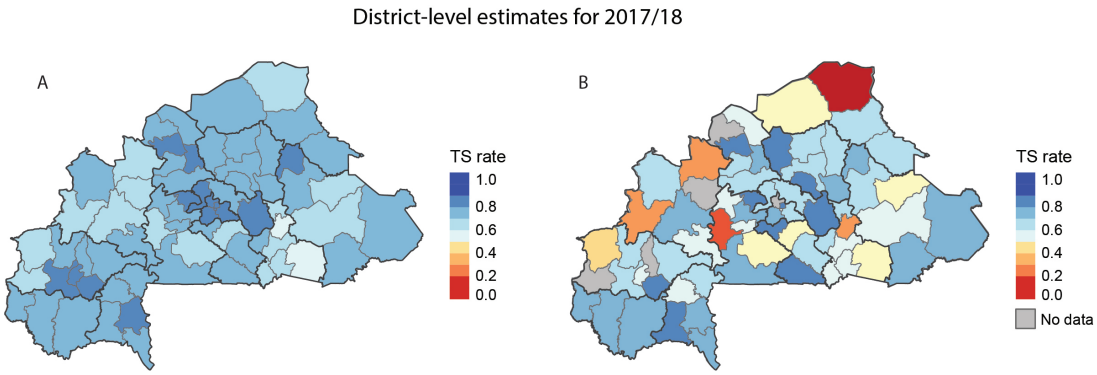


Figure A.10. District-level aggregated surface for 2017/18. A) District-level treatment-seeking (TS) rate aggregated using the pixel-level estimates from the model. B) District-level estimates from the 2017/18 MIS. Survey is not powered at the district level, so some districts have no estimated data.

APPENDIX B

Supplement to: Chapter 3

B.1. Map of health districts in Burkina Faso

Figure B.1. Map of the 70 health districts in Burkina Faso with names. Colors correspond to the SMC rollout group to which each district belongs.

B.2. Data imputation and aggregation results

Table B.1 shows a summary of the results from the imputation processes outlined in the methods section of the main text. All-cause outpatient visits were missing at the same rate for both age-groups and the imputation procedure employed affected them equally. Confirmed malaria cases were missing

Imputed column	Age group	% missing originally	% missing imputed	# rows imputed	% missing remaining	# HFs with imputed values
confirmed malaria	under 5	33.1%	4.1%	1,921	31.7%	983
	over 5	31.8%	4.1%	1,842	30.5%	986
cases all-cause outpatient	under 5	20.2%	2.9%	846	19.6%	488
	over 5	20.2%	2.9%	846	19.6%	488
cases malaria hospitalizations	under 5	61.2%	0.8%	743	61.4%	244
	over 5	58.0%	1.2%	983	57.3%	333

Table B.1. Summary table of data completeness before and after imputation. For each column in the data imputed and each age group, we show the percentage of rows which were originally missing, the percentage and number of those missing observations successfully imputed, and the percentage of rows still missing after the imputation procedure. We also show the number of unique health facilities (HFs) which had values imputed.

at a larger rate among children u5, however, both age groups had a similar percent of missing values imputed. Malaria hospitalizations were missing over 50% of the time for the ov5 age group and almost two-thirds of the time for the u5 age-group. The imputation procedure employed did not allow for substantial imputation on these columns due to the conservative approach employed, meaning that most health facilities from which there were missing values either had too many, or too many consecutive, missing values or in the time series (Table B.1).

Cleaning the data resulted in the removal of over two-thirds of health facilities from the data modeling malaria incidence or malaria proportion of outpatient visits and the removal of over 90% of health facilities from the data modeling malaria hospitalizations, respectively. Requiring complete reporting guarantees non-missing counts of confirmed malaria cases, all-cause outpatient visits, and malaria hospitalizations among both age-groups for the 16 months of SMC rollout (Table B.2). This step also removed one district, Pouytenga from the 2016 SMC group from the data modeling malaria incidence or malaria proportion of outpatient visits and the data modeling, as none of the reports for the 16 modeling months were complete and consistent. In addition, 9 districts were removed from the data modeling malaria hospitalizations: 8—Dafra, Gorom-Gorom, Koudougou, Manga, Pouytenga, Sapone, Thiou, and Yako—from the 2016 SMC group and 1—Lena—from the 2017. District-level time series for the three indicators analyzed across the two age groups are shown in supplement Figure B.2.

Variable analyzed	% of rows removed	# rows remaining	# HFs remaining	# districts remaining
Malaria incidence & malaria proportion of outpatient visits	65.4%	37,920	790	50
Malaria hospitalizations	91.1%	9,744	203	42

Table B.2. Summary table of data cleaning process. For both datasets produced from the cleaning process, we show the percentage of rows which were removed according to the inclusion criteria, and the number of rows, unique health facilities (HFs), and districts remaining in the data after the process.

B.3. Time series of Burkina Faso HMIS data used for modeling

We plot the time series for 6 of the malaria indicator variables used for modeling among the districts used for modeling. Only months in which SMC was distributed each year for modeling. For every year in which SMC was distributed to these districts, the model compares the change in the data from the pre-SMC to the post-SMC time periods in districts which received SMC for the first time versus districts which have not yet received SMC. Model 1 is used to estimate the effect of SMC on reducing the malaria burden using these 6 indicators as the response variable.

B.4. Model diagnostics

Below, we present the model diagnostics for model 1 applied to the malaria proportion of outpatient visits among children under-5. Similar model behavior is seen for all malaria indicators used for model 1 and for model 2 (see Chapter 3 methods).

The diagnostic plots shown below show the behavior of the model under the GLMM assumptions. Figure B.3 shows evidence of few outliers in districts like Diebougou and Dano by the departure of the plotted points from the horizontal black line. We also see that the model is fitting the data well by the proximity of the points to the diagonal black line in Figure B.4. Figure B.5 corroborates this, although we see some departure from the diagonal line in districts like Diebougou and Dano.

Figure B.6 shows the Q-Q plot for the standardized residuals and signal no strong violation of the assumption of a Poisson distribution in our model, as evidenced by the points falling in a diagonal line

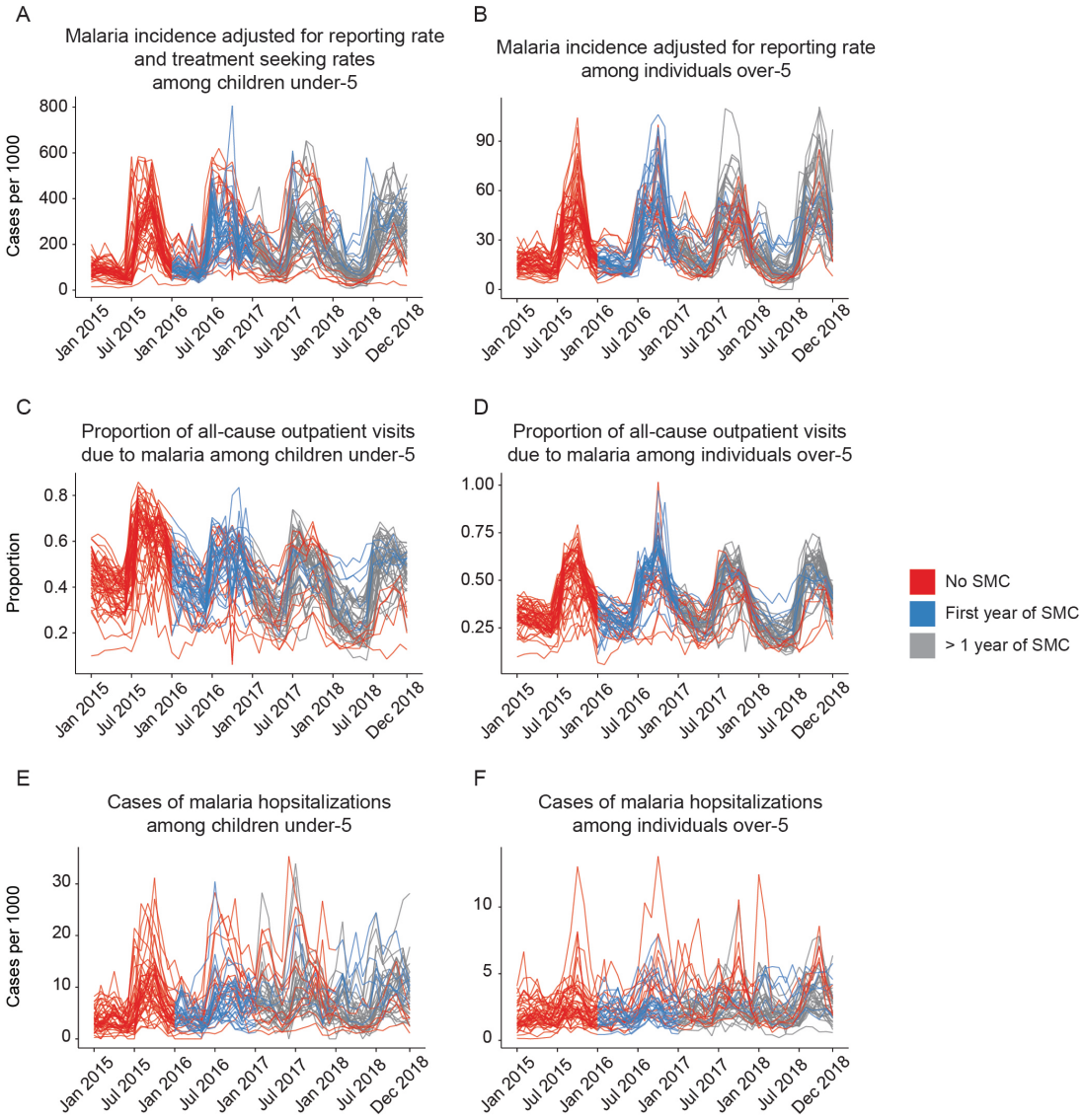


Figure B.2. Time series of Burkina Faso HMIS data used for modeling. Each line represents one health district with the color of the line representing the state of the SMC program in that district. A) Confirmed malaria cases from RDT or microscopy among children under-5 adjusted for health facility reporting rate and changes in febrile treatment seeking from 50 health districts. B) Confirmed malaria cases from RDT or microscopy among individuals over-5 adjusted for health facility reporting rate from 50 health districts. C) The proportion of all-cause outpatient visits due to malaria among children under-5 from 50 health districts. D) The proportion of all-cause outpatient visits due to malaria among individuals over-5 from 50 health districts. E) Number of hospitalized malaria cases adjusted for health facility reporting rate among children under-5 from 42 health districts. F) Number of hospitalized malaria cases adjusted for health facility reporting rate among individuals over-5 from 42 health districts.

across all districts. Again, only Diebougou and Dano raise some concerns. Finally, Figure B.7 shows the Q-Q plot for the random effects and, while some points do fall outside the diagonal line, we determine no strong violation of the assumptions of normality placed on the random effects. Similar behavior is seen for all malaria indicators used for model 1 and for model 2.

The model assumes a compound-symmetric variance-covariance structure and shows a significant negative correlation of -0.81 (95% CI: [-0.93, -0.54]) between the random effect regression coefficients of lag1 and lag2 precipitation. Similar significant negative correlations, ranging between -0.94 and -0.76, are seen between these precipitation terms for all malaria indicators used for model 1 and for model 2.

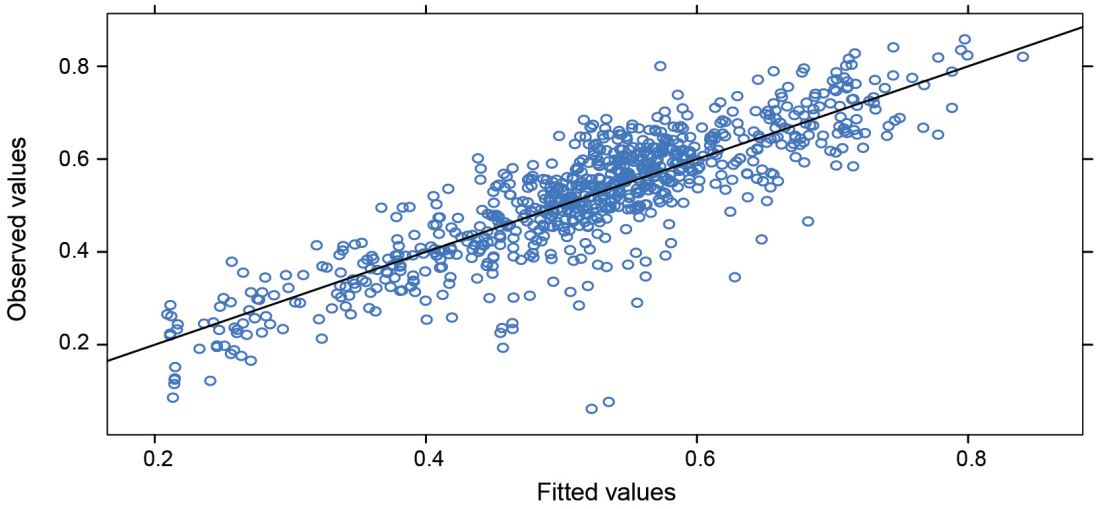


Figure B.3. Scatterplot of standardized residuals against fitted values for each of the 50 districts. Horizontal line in black represents the y=0 line. Diebougou and Dano show evidence of outlier observations.

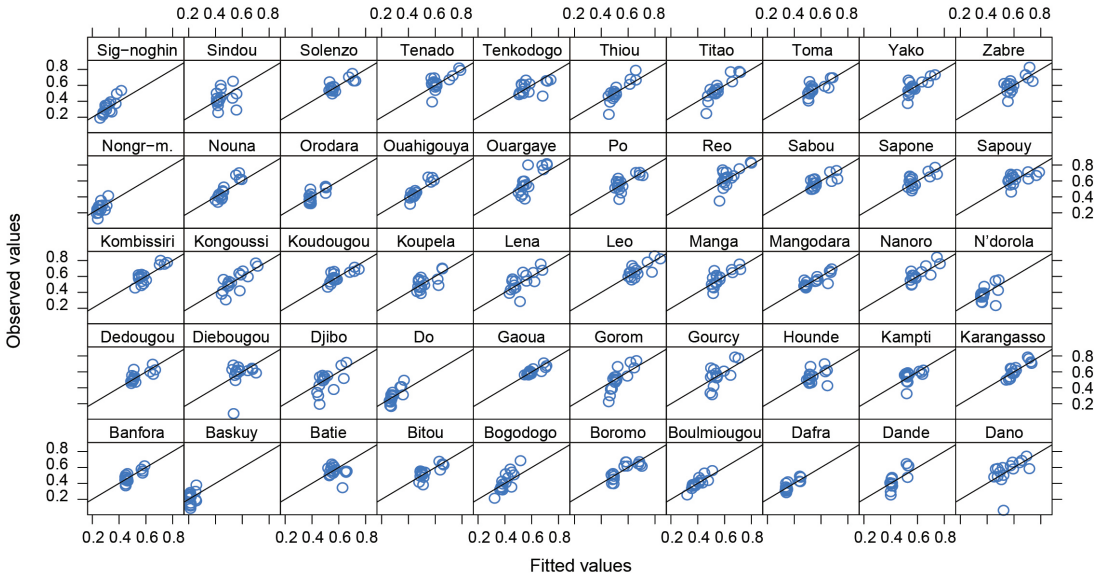


Figure B.4. Scatterplot of observed against fitted values among all 50 health districts modeled. Diagonal line in black represents the y=x line. The proximity of the points to the diagonal shows that the appropriate fit of the model.

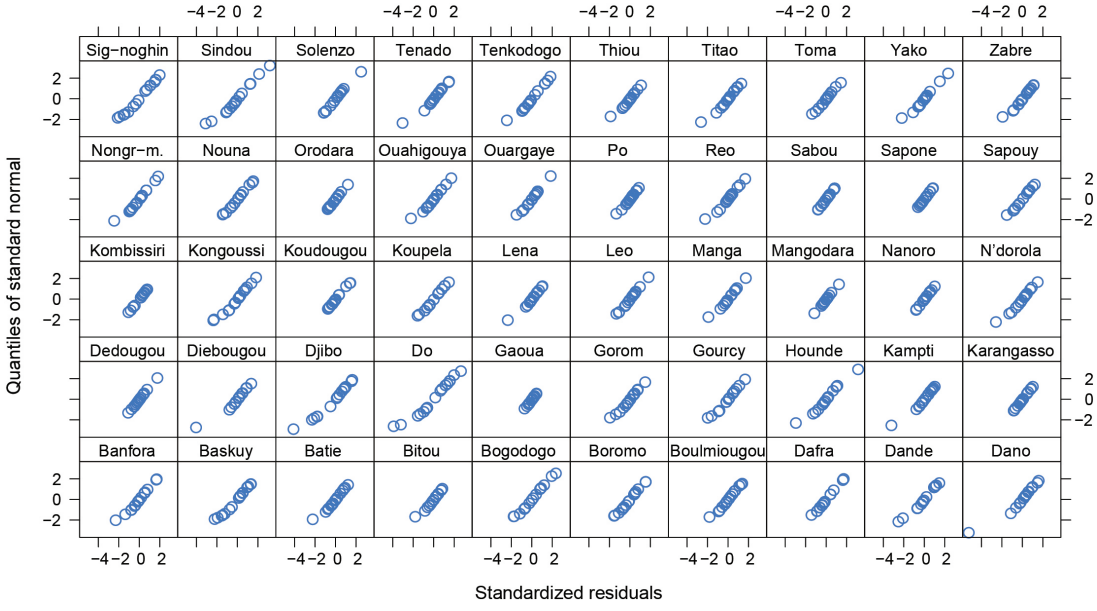


Figure B.5. Scatterplot of observed against fitted values for each of the 50 health districts modeled. Each diagonal line in black represents the y=x line. The district level plot shows some outlying observations in the Diebougou and Dano districts.

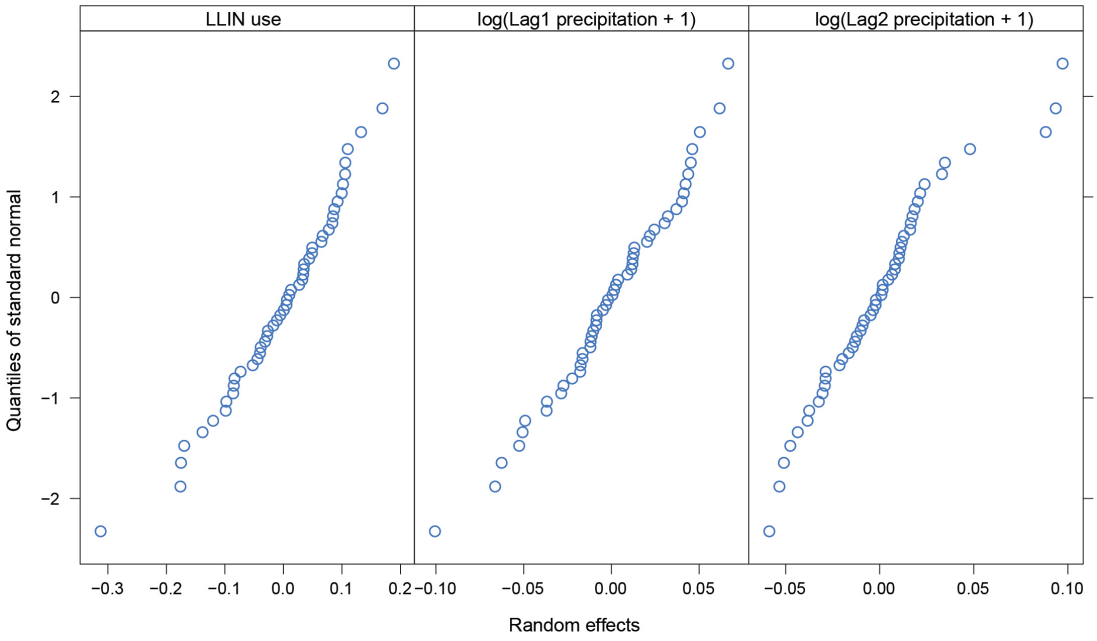


Figure B.6. Q-Q plot of standardized residuals for all 50 health districts modeled. No district shows a strong violation of the assumption of a Poisson distribution in the model. Diebougou and Dano show outlying observations.

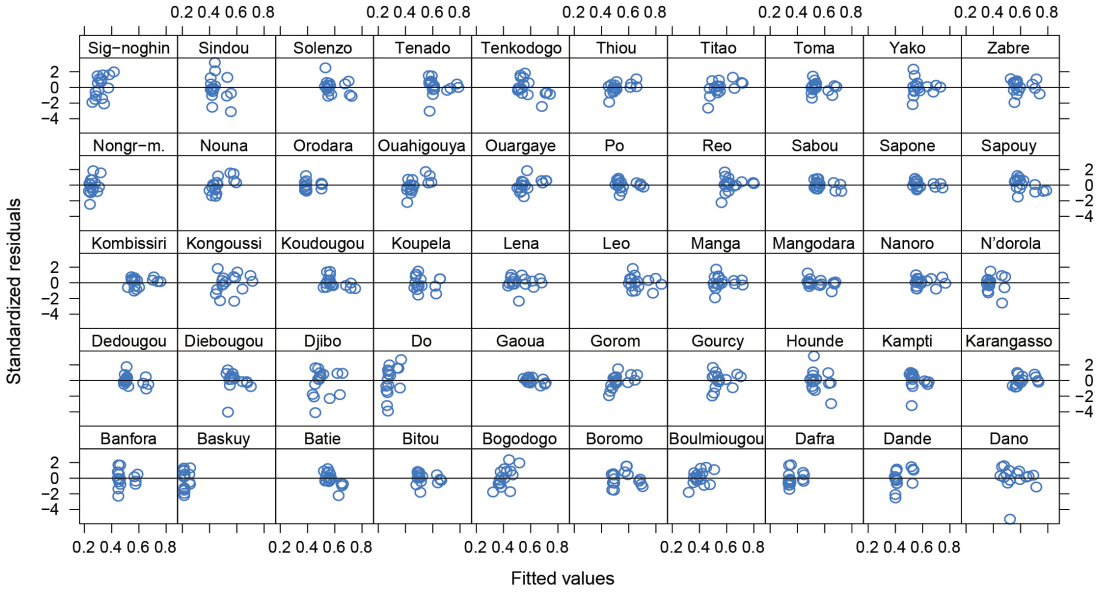


Figure B.7. Q-Q plot of standardized residuals for the three random effects terms used. While some observations do not fall along the diagonal, these plots show no strong violation of the normal assumptions.

APPENDIX C

Supplement to: Chapter 4

C.1. Model selection & diagnostics

We assess model performance for the inclusion of the latent treatment-seeking rate as a predictor variable and inclusion of climate covariates. Model selection for was done in a forward step-wise framework, where models were considered by adding one covariate at a time. Climate covariates considered are monthly precipitation and monthly surface air temperature collected at a 30 km grid mesh (Hersbach et al., 2020). Covariates were assessed based on significance of the predictor variable and we assess model performance via the deviance information criterion (DIC) and Watanabe–Akaike information criterion (WAIC) (Ingebrigtsen et al., 2014).

All models presented below include the fixed effects corresponding to the difference-in-differences framework to measure the effect size of SMC. They also include both structured random effects for the seasonal temporal component and the spatio-temporal component which has a Matérn covariance structure. (Chapter 4, Equation 4.2.2).

We assess the performance of the latent process for the modeled treatment-seeking rate at the 1×1 km scale (TS rate) and at the aggregated district level (TS rate DS). Both of these covariates show to be significant in the statistical models, with the latent process included at the 1×1 km scale producing the best performing model according to the DIC and WAIC values (Table C.1). This predictor is retained for the subsequent models which explore the inclusion of climate covariates.

Table C.1 shows the models considered that include the significant climate covariates. Focusing on monthly precipitation, we explore the inclusion of the significant precipitation covariates, which include precipitation of the current month and the previous two months (lag-0, lag-1, and lag-2). The precipitation of the current month (lag-0) produces the best performing model according to DIC and WAIC values. For

Model covariates	DIC	WAIC
TS rate	350212.3	368442.1
TS rate DS	350307.8	368675.6
TS rate Precipitation	350236.8	368514.9
TS rate Precipitation lag1	350311.6	368717.4
TS rate Precipitation lag2	350427.7	369007.6
TS rate Precipitation Precipitation lag1	350308.8	368711
TS rate Precipitation lag1 Precipitation lag2	350455.8	369085.9
TS rate Precipitation Precipitation lag1 Precipitation lag2	350361.2	368845
TS rate Air temperature	350249.2	368539.6
TS rate Air temperature lag1	350254.3	368551.9
TS rate Air temperature Air temperature lag1	350249.1	368539.4
TS rate Precipitation Air temperature	350237.3	368516.1
TS rate Precipitation lag1 Air temperature lag1	350311.6	368716.8
TS rate Precipitation Air temperature Precipitation lag1 Air temperature lag1	350323.6	368748.4
TS rate Precipitation lag1 Precipitation lag2 Air temperature lag1	350345.3	368807.1

Table C.1. Table for predictor covariates considered in the healthcare facility-level model of the malaria burden. Covariates considered include the modeled latent process for fine-scale treatment-seeking rates and the climate covariates. Model selection is done in a step-wise manner, considering the final model by minimizing the DIC and WAIC.

monthly surface air temperature, we explore the inclusion of surface air temperature of the current and previous month (lag-0 and lag-1). The surface air temperature of 2 months prior (lag-2) proved not to be significant in the model and thus was not considered in table C.1. Including both surface air temperature values (lag-0 and lag-1) produced the best model according to DIC and WAIC metrics, marginally beating out the model with just the surface air temperature of the current month. Considering the ensemble of both climate covariates (precipitation and surface air temperature), we see no improvements in the model. Ultimately, the inclusion of the climate covariates does not improve model performance. Thus we retain the model with just the treatment-seeking rate at a 1×1 km scale for the final model.

C.2. Derivation for difference-in-difference effectiveness for SMC

Here, we present the derivation of the effect size of SMC in reducing the malaria burden using the model in Chapter 4 (Equation 4.2.2). The model formulation is simplified by assuming that the temporal and spatio-temporal random effects ρ_t and $\omega_{t,h}$ are fixed to their zero mean values.

The simple formulation of the model is then written as follows:

$$log(\mathrm{E}[y_{m,y,h}]) = \alpha_y + \gamma_{g(h)} + \delta_{m,y,g(h)} + \beta Z_{m,y,h}$$

where α_y is the year fixed effect for $2015 - 2018$, $\gamma_{g(h)}$ is the SMC group fixed effect, where $g(h)$ defines the SMC group health facility h belongs to, and $g(h) \in \{2016, 2017, 2018\}$. $\delta_{m,y,g(h)}$ is the DiD effect term, where $\delta_{m,y,g(h)} = 1$ if $y \geq g(h)$ and $m \geq 7$. $Z_{m,y,h}$ is the modeled treatment-seeking rate. The treatment-seeking rate follows a step-function change, where the values have a constant rate prior to June 2016 and changes to a different rate after this date.

We write the model as separate dummy variables and indicator functions. To ease the notation, we will enumerate dates (m, y) as $t \in \{1, 2, \ldots, 48\}$. $Z_{t,h}$ will be written as Z_h prior to June 2016 ($t = 18$) and Z_h^* after June 2016. t_1 represent any time point t prior to the start of the intervention, and t_0 represents any time point t before the start of the intervention. The model is re-written as:

$$log(\mathrm{E}[y_{t,h}]) = \alpha_{2015} I[t \leq 12] + \alpha_{2016} I[13 \leq t \leq 24] + \alpha_{2017} I[25 \leq t \leq 36] + \alpha_{2018} I[37 \leq t \leq 48] +$$
$$\gamma_{2016} I[g(h) = \text{“2016”}] + \gamma_{2017} I[g(h) = \text{“2017”}] + \gamma_{2018} I[g(h) = \text{“2018”}] +$$
$$\gamma_{2019} I[g(h) = \text{“2019”}] + \delta_{2016} I[g(h) = \text{“2016”}] I[t \geq 19] +$$
$$\delta_{2017} I[g(h) = \text{“2017”}] I[t \geq 31] + \delta_{2018} I[g(h) = \text{“2018”}] I[t \geq 42] +$$
$$\beta Z_{t,h}$$

We will show the derivation for the effect size for 2016 group. This same derivation may be extended for the 2017 and 2018 groups.

C.2.1. First difference in treatment group

We first derive the difference in the treatment group pre- and post-treatment. This derivation may be extended to estimate the effect for the 2017 and 2018 rollout groups. The first difference is then,

$$log(\mathrm{E}[y_{t,h}|19 \leq t < 31, g(h) = \text{“2016”}]) - log(\mathrm{E}[y_{t,h}|t < 19, g(h) = \text{“2016”}])$$

where

$$log(\mathrm{E}[y_{t,h}|19 \leq t < 31, g(h) = \text{“2016”}]) = \alpha_{[y(t) \geq 2016]} + \gamma_{2016} + \delta_{2016} + \beta Z^*_{h_T}$$
$$log(\mathrm{E}[y_{t,h}|t < 19, g(h) = \text{“2016”}]) = \alpha_{[y(t) \leq 2016]} + \gamma_{2016} + \beta(I[t < 18] Z_{h_T} + I[t \geq 18] Z^*_{h_T})$$

If we use $t = 18$ as the pre-treatment period, we get that the first difference is:

$$\alpha_{[y(t) \geq 2016]} - \alpha_{[y(t) \leq 2016]} + \delta_{2016}$$

If we use $t < 18$ as the pre-treatment period, we see that the first difference has a lingering term related to the treatment-seeking rate:

$$\alpha_{[y(t)\geq 2016]} - \alpha_{[y(t)\leq 2016]} + \delta_{2016} + \beta(Z^*_{h_T} - Z_{h_T})$$

C.2.2. Second difference in control group

We now derive the difference in the control group pre- and post-treatment. The second difference is then,

$$log(\mathrm{E}[y_{t,h}|19 \leq t < 31, g(h) \neq \text{“2016”}]) - log(\mathrm{E}[y_{t,h}|t < 19, g(h) \neq \text{“2016”}])$$

where

$$log(\mathrm{E}[y_{t,h}|19 \leq t < 31, g(h) \neq \text{“2016”}]) = \alpha_{[y(t)\geq 2016]} + \gamma_{g(h)} + \beta Z^*_{h_C}$$

$$log(\mathrm{E}[y_{t,h}|t < 19, g(h) \neq \text{“2016”}]) = \alpha_{[y(t)\leq 2016]} + \gamma_{g(h)} + \gamma_{2016} + \beta(I[t < 18]Z_{h_C} + I[t \geq 18]Z^*_{h_C})$$

If we use $t = 18$ as the pre-treatment period, we get that the first difference is:

$$\alpha_{[y(t)\geq 2016]} - \alpha_{[y(t)\leq 2016]}$$

If we use $t < 18$ as the pre-treatment period, we see that the first difference has a lingering term related to the treatment-seeking rate:

$$\alpha_{[y(t)\geq 2016]} - \alpha_{[y(t)\leq 2016]} + \beta(Z^*_{h_C} - Z_{h_C})$$

C.2.3. Diff-in-Diff for 2016 effect

The diff-in-diff effect for 2016 if we condition on comparing $t < 18$ as the pre-treatment time period is:

$$\textbf{Diff 1} - \textbf{Diff 2} = \delta_{2016} + \beta\big((Z^*_{h_T} - Z_{h_T}) - (Z^*_{h_C} - Z_{h_C})\big)$$

The diff-in-diff effect for 2016 if we condition on comparing $t \geq 18$ as the pre-treatment time period is:

$$\textbf{Diff 1} - \textbf{Diff 2} = \delta_{2016}$$

Thus, δ_{2016} represented the effect size for the 2016 SMC rollout group, while the actual difference-in-differences between the intervention and control group includes both the effectiveness of the SMC intervention and the lingering effect of the change in treatment-seeking rate.

C.3. Model hyperparameters for random effects

We present the hyperparameters for the healthcare facility-level model, capturing the malaria proportion of outpatient visits (Chapter 4, Equation 4.2.2) in an integrated nested Laplacian approximation (INLA) framework with a generalized linear mixed-effects modeling (GLMM) approach. The hyperparameters are estimated for the two random effects in the model, representing the temporal random effect with a seasonal correlation structure and the spatio-temporal random effect with a Matérn spatial correlation structure and an i.i.d. temporal correlation (Table C.2).

For the seasonal random effect, the vector of random effects $\mathbf{u} = (u_1, \ldots, u_n)$ is assumed to have periodicity p and whose components are defined as the sums $x_i + x_{i+1} + \cdots + x_{i+p-1}$ which are assumed to be independent Gaussians with zero mean and precision τ. The precision term τ modifies the precision matrix of the model $Q = \tau R$, with R having the neighborhood structure of the model (Gómez-Rubio, 2020). The hyperparameter precision term τ then represents the *reciprocal* of a variance parameter.

For the spatio-temporal random effect, there are only the hyperparameters for the spatial structure of the random effect corresponding to a Matérn correlation structure, as the i.i.d. term has no hyperparameters. The Matérn structure is parametrized by a *range* parameter and the standard deviation. Since the spatial domain is defined in the decimal degree coordinate scale, the values will appear small by

nature since we are estimating the correlation of healthcare facilities within one country. The range parameter describes the distance at which the correlation decays past a certain threshold, commonly below a correlation of 0.05 (Lindgren & Rue, 2015). A range parameter of 0.018 implies that the correlation of the spatial effect goes below 0.05 after approximately 2 km, which makes sense due to the clustering of a large amount of healthcare facilities in certain regions (mainly the Centre region) of the country (Figure 4.1).

Model hyperparameters	**mean**	**95% CI**
Precision for time effect	29.32	[23.85, 36.45]
Range for spatial field	0.018	[0.016, 0.020]
St.dev for spatial field	2.06	[1.89, 2.27]

Brackets indicate 95% CI.

Table C.2. Model hyperparameters for Chapter 4, Equation 4.2.2. Precision term is shown for temporal seasonal random effect. Range and Standard Deviation are shown for "spatial field", which corresponds to the spatio-temporal random effect with Matérn spatial correlation structure. The spatial field is in the scale of coordinate degrees (decimal degrees).

C.4. Maps of complete time series

Here, we show the maps of the estimated malaria proportion of outpatient visits at a 1×1 km resolution for the entire 48 months in the study period. As mentioned in the Methods section of Chapter 4, these maps represent smoothed estimates of the malaria proportion of outpatient visits calculated with Equation 4.2.2 and removing the spatial random effect $\omega_{t,h}$.

The 48 maps are shown below in Figure C.1 and Figure C.2. As in Figure 4.2, districts which receive SMC for the first time each year are outlined in red (Figure C.2). The seasonality of the estimated malaria burden is seen in the figures, with malaria peaking in August – December and with the low burden valley in February – June.

Scanning across each row in the figures, we can see the change in the malaria burden for a fixed month across the years. The reductions in the malaria proportion of outpatient visits are clearly seen for all months, with the exception of a slight increase between January through June of 2015 and 2016, which are time periods prior to SMC being introduced in any of the SMC rollout groups analyzed (Figure

C.1). Reductions are seen most clearly in districts which have initiated the SMC campaign, (outlined in red, Figure C.2). A slight increase is seen in the final months of 2018, especially in the 2016 SMC rollout group, which signals a reversal of the decreases in the malaria burden attained with the introduction of SMC.

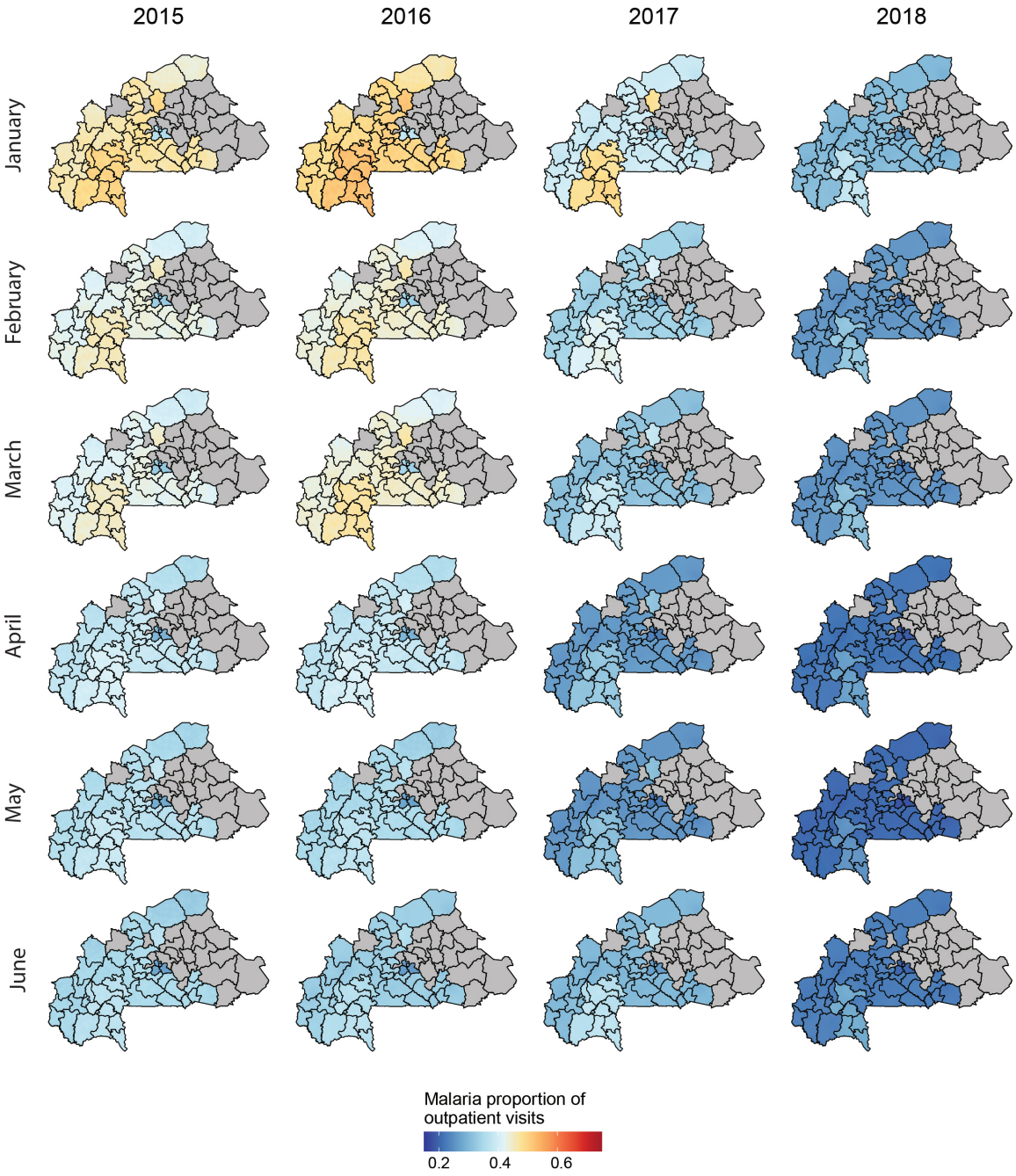


Figure C.1. Maps of modeled estimates for the malaria proportion of outpatient visits at a 1 x 1 km scale. The months mapped correspond to the months of SMC distribution. Districts are outlined in each map and districts with grey color indicate districts which were not included in the model.

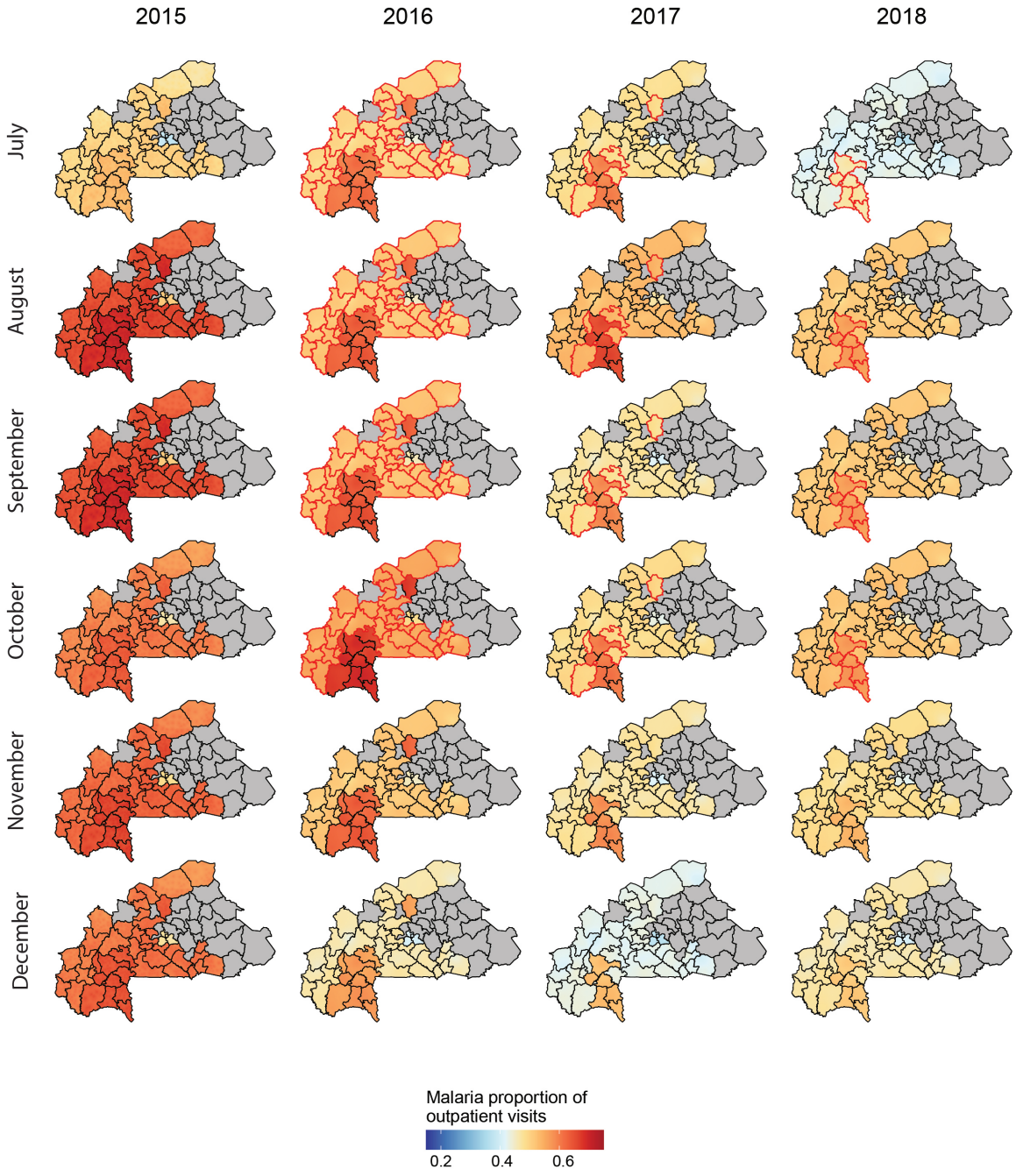


Figure C.2. Maps of modeled estimates for the malaria proportion of outpatient visits at a 1 x 1 km scale. Districts are outlined in each map, and red outlines indicate districts which initied SMC in that year. Districts with grey color indicate districts which were not included in the model.