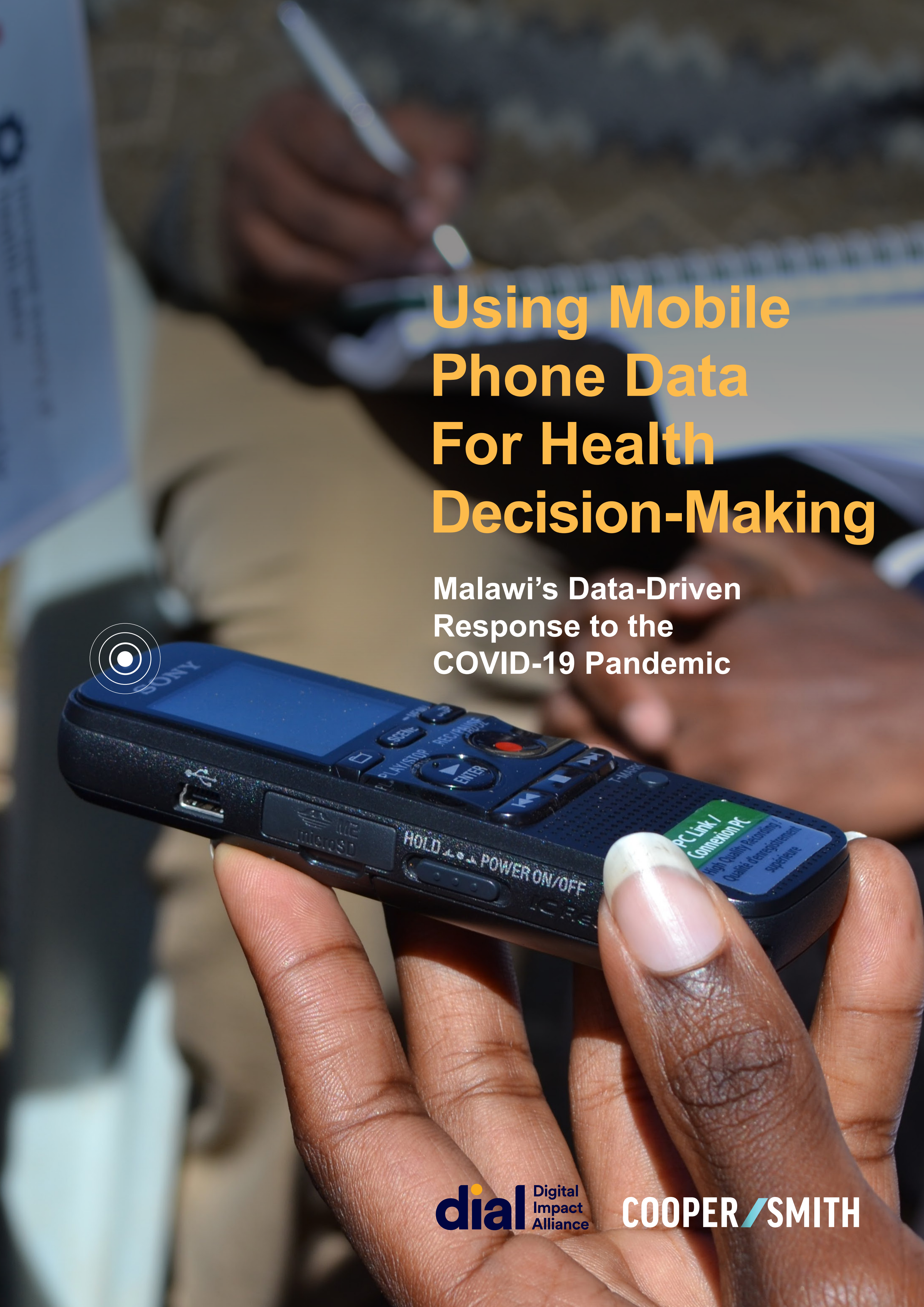


Using Mobile Phone Data For Health Decision-Making

Malawi's Data-Driven Response to the COVID-19 Pandemic

dial Digital Impact Alliance

COOPER/SMITH

ACKNOWLEDGMENTS

This technical report was prepared as part of a partnership between the Digital Impact Alliance (DIAL) at the United Nations Foundation (UNF) and Cooper/Smith. Primary contributing authors include Dylan Green, Michael Moszczynski, Cassie Morgan, Guillaume Foutry, Rachel Sibande, and Tyler Smith. Special thanks are due to other members of the UNF/DIAL team, as well as partner organizations that supported various investments relevant to this engagement, including the Bill & Melinda Gates Foundation (BMGF), the United Nations Children's Fund (UNICEF), the Laboratório de Imunopatologia Keizo Asami (LIKA), the United Nations Development Programme (UNDP), and the United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA) for data and technical support.

Cooper/Smith would expressly like to thank the Tableau Foundation for providing pro bono access to a suite of products and making advanced data visualization tools easily accessible to stakeholders in Malawi. Cooper/Smith is also grateful for the feedback and comments provided by the Development Impact Evaluation (DIME) Department at the World Bank, which helped refine the population mobility model.

dial Digital Impact Alliance

ABOUT DIAL

The Digital Impact Alliance is an independent global alliance funded through a collaboration among the Bill & Melinda Gates Foundation; the UK's Foreign, Commonwealth & Development Office (FCDO); and the Swedish International Development Cooperation Agency (Sida), and hosted within the UN Foundation in Washington, D.C. DIAL was established in 2015 as a "think, do, replicate" tank. We combine practical research with evidence-based advocacy to advance digital inclusion to achieve the Sustainable Development Goals (SDGs). DIAL identifies barriers to the routine use of digital solutions and data by development actors (countries, NGOs, multilateral institutions), tests ways to remove them, and packages solutions for these actors to use in service delivery efforts.
www.digitalimpactalliance.org.

COOPER/SMITH

ABOUT COOPER/SMITH

Cooper/Smith is a technical assistance organization that uses hard data to increase the effectiveness and efficiency of development programs worldwide. Cooper/Smith supports large-scale data systems and data use projects in more than a dozen countries throughout the world. Our team of technical specialists brings extensive experience in information communication technology for development (ICT4D), systems integration, software engineering, data use analyses, results tracking and monitoring, and high-frequency data collection, particularly in the domains of health, supply chain, and resilience. Central to our approach is a focus on sustainability and empowering government, country partners, and local organizations with the data, capacity, and resources required to support data-driven solutions. As an international organization, Cooper/Smith participates in numerous technical and advisory groups and is committed to the democratization of data access and use worldwide. For more information, please visit our website at www.coopersmith.org.

UNITY AND FREEDOM

ABOUT THE MINISTRY OF HEALTH AND POPULATION OF MALAWI

The Mission of the Ministry of Health and Population of Malawi is to provide strategic leadership for the delivery of a comprehensive range of quality, accessible, and efficient health services to all Malawians through the creation and sustenance of a strong health system.
www.health.gov.mw

TABLE OF CONTENTS

ABBREVIATIONS

BMGF	Bill & Melinda Gates Foundation
CDR	Call detail record
CIP	Capital Investment Plan
DIAL	Digital Impact Alliance
DIME	Development Impact Evaluation
ECOWAS	Economic Community of West African States
EOC	Emergency Operation Center
FCDO	Foreign, Commonwealth & Development Office
GDPR	General Data Protection Regulation
GoM	Government of Malawi
GSMA	Global System for Mobile Communications Association
ICU	Intensive care unit
IRB	Institutional Review Board
LIKA	Laboratório de Imunopatologia Keizo Asami (Keizo Asami Immunology Lab)
MACRA	Malawi Communications Regulatory Authority
MNO	Mobile network operator
MoH	Ministry of health
NCST	National Commission for Science & Technology
NGO	Nongovernmental organization
NPC	National Planning Commission
NSO	National Statistical Office
PHIM	Public Health Institute of Malawi
SEIR	Susceptible-exposed-infected-recovered
SIDA	Swedish International Development Cooperation Agency
TA	Traditional authority
UNDP	United Nations Development Program
UNF	United Nations Foundation
UNOCHA	United Nations Office for the Coordination of Humanitarian Affairs
USAID	United States Agency for International Development
VPC	Virtual Private Cloud
WHO	World Health Organization
WAHO	West African Health Organization

EXECUTIVE SUMMARY

In 2017, the Digital Impact Alliance and Cooper/Smith embarked on a collaborative effort with the Malawi Ministry of Health (MoH) and the Malawi Communications Regulatory Authority (MACRA) to demonstrate how mobile network operator (MNO) data and analytics can be used to improve decision-making within Malawi's national health system.

Background: Phase I (2017-2019)

In 2017, the MoH launched a Capital Investment Plan (CIP), including a proposal to build 900 new health posts across the country. One of the primary objectives of the CIP was to strategically identify locations for these new health posts to ensure that 95% of the Malawian population would live no more than five to six kilometers from a health facility by 2023.

To help the ministry accomplish this objective, Cooper/Smith developed an analytics model utilizing MNO and static datasets (i.e., fixed data that does not change after being recorded) that generated dynamic maps (i.e., reflecting changes according to changes in the data) with optimal site locations for all new health posts across Malawi. The model helped the MoH understand population growth, density, and migration at a granular level, enabling the country to make data-driven decisions about the optimal placement of new health facilities in line with Malawi's national priorities. As a result, the ministry is now equipped with the data and tools needed to make informed decisions about the next stage of primary health care (PHC) in Malawi. In 2017, 55.3% of Malawians lived no more than five kilometers from a health facility. By using the results of the model to optimize resource allocation and decide where to build the next 900 facilities, this number could be brought to 95% by 2023.

Supporting the COVID-19 Response (2020-2021)

In early January 2020, COVID-19 was declared a Public Health Emergency of International Concern by the World Health Organization (WHO). In March of 2020, the MoH asked Cooper/Smith to pivot the methodology and learnings from Phase I to support the country's national COVID-19 response efforts. The ministry first requested support developing epidemiological estimations and digital tools to aid in the detection of population patterns that could result in high-risk events for COVID-19 infection. With support from DIAL, Cooper/Smith then designed a strategy to produce analytic and response tools that could quickly address the country's urgent needs by leveraging Malawi's mobile network operator data.

Methodology and Results

The main data source Cooper/Smith analyzed was a list of anonymized call detail records (CDRs) provided by one of Malawi's two MNOs, representing a set of outgoing voice calls or SMS messages sent from an individual cell tower. Cooper/Smith used a clustering algorithm to reduce more than 10,000 unique cell tower IDs to just 729 clusters. While MNO data is very carefully and completely anonymized, making it impossible to pinpoint the exact location of an individual, Cooper/Smith was able to calculate a coverage area for each cluster, generating a map to compare with the traditional authorities (TAs) of Malawi and assessing population density and mobility within and between clusters.

To validate the use of anonymized CDR data as a proxy for population, Cooper/Smith proved the correlation by using actual population figures. This was achieved by analyzing the overlapping areas between clusters against the TAs of Malawi, thereby calculating the predicted subscriber population of each TA. Predictions from the model were similar to the UN population predictions for 2020, which proved the viability of developing a population mobility model based on phone subscribers' data. In addition, Cooper/Smith addressed MNO data biases, particularly gender and income gaps, by limiting the model's use to relative movement, e.g., an event was defined in comparison to previous data for that location.

To validate the use of anonymized CDR data as a proxy for population, Cooper/Smith proved the correlation by using actual population figures.

Several other data sources were used to inform the models:

- The GPS coordinates of all the cell towers provided by the MNO
- A country map of the population density estimates provided by WorldPop, a project of the University of Southampton
- A shapefile of Malawi's administrative divisions and UN estimates of population growth since the last census, which provided a grid of 1km² tiles and estimates how many people live in each tile

In parallel, Cooper/Smith also developed a dynamic, compartmental, deterministic susceptible-exposed-infected-recovered (SEIR) model,[2] which acted as a mathematical model of COVID-19 transmission.

Working jointly with stakeholders from the MoH, the Public Health Institute of Malawi (PHIM), and the national COVID-19 Emergency Operation Centre (EOC), Cooper/Smith designed two use cases that addressed the primary needs of the MoH:

Use Case 1: Simulate the spread of the pandemic across the cellular network to improve epidemiological estimations. Cooper/Smith combined the SEIR model's projections with population mobility insights derived from the CDR data analysis to estimate the epidemic start date in any specific TA. This was done to refine the model and improve its accuracy. Cooper/Smith leveraged estimates of interchange (the fraction of mobile subscribers in one TA that also visits another), or "connectedness" between TAs, from the CDR data.

From here, Cooper/Smith could identify when what was believed to be the first (or nth) infection of each TA likely took place. It integrated the refined SEIR projections into a dashboard for MoH stakeholders, enabling them to visualize potential epidemic scenarios, which were updated daily.

---

[1] WorldPop Age and Structures / Individual Countries: https://www.worldpop.org/doi/10.5258/SOTON/WP00538
[2] The model is available free of charge on GitHub: https://github.com/CooperSmith-org/Malawi-COVID-Epidemiological-Model
[3] Denotes an unspecified item or instance in a series, typically the last or latest in a long series: "For the nth time that day, they were forced to relate the whole story."

Use Case 2: Detect anomalous mass gatherings to identify potential transmission events and tailor public health response locally. Mass event detection, namely events that had occurred the previous day that showed abnormally high numbers of subscribers gathering in one place, was key to Malawi's COVID-19 response. The MNO data provided not only users' locations but their movements throughout the day as well. Therefore, if a mass event represented a potential locus of infection, Cooper/Smith could use the same data to determine where the risk had, or was likely to, spread.

This made it possible to generate a "risk score" for areas surrounding the event, distinguishing between events that were mostly regional and events that attracted subscribers from other parts of the country. All information on identified mass events and the associated risk were accessible for MoH decision-makers though dashboards that were updated daily with data from the day prior.

Assuring Ethical Responsibility

The novelty of working with data and computational sciences in the development field, coupled with increasing surveillance concerns, has heightened the need for stakeholders to prioritize compliance with best practices in data management to mitigate concerns over privacy and ethics. Responsible data use, protection of privacy, and strict avoidance of unintended uses of data are essential to ensure trust and proper use, as well as protect individual and group privacy.

Before the MNO shared data with Cooper/Smith, it preprocessed the data to ensure privacy protection and quality assurance. The MNO did this by:

- Utilizing an MD5 hashing algorithm[4] that stripped all identifying data—such as name, phone number, gender, and age—and assigned each user an individual code unrelated to the individual's phone number
- Removing the timestamp and replacing it with a four-hour block, which prevented potential triangulation of a user's location and movements

In accordance with national and global best practices related to data regulations and security, Cooper/Smith obtained authorization from the Institutional Review Board (IRB) of the National Commission for Science & Technology (NCST) of Malawi to perform a mobile data analysis, using anonymized CDR data sets provided by the MNO and authorized by MACRA. Additionally, Cooper/Smith and an MNO based in Malawi signed a data sharing agreement (DSA) that established a framework and mechanisms on what data is shared between project partners. Contracts for Collaboration provides examples of DSA agreements between MNOs and third-party organizations. Additionally, Cooper/Smith created an MNO data processing guide document detailing data pipeline, hosting infrastructure, and outputs.

To ensure compliance with best practices regarding the use of sensitive data and the design of algorithms and models for public health purposes, Cooper/Smith completed a risk/benefit self-assessment reviewed by the DIAL compliance team and an external assessment led by the United Nations Office for the Coordination of Humanitarian Affairs. Both assessments validated the methodologies used by the project against global standards in data privacy and security.

---

[4] Anonymizing phone numbers is done using an MD5 hashing algorithm and a salt value. This involves appending an arbitrary string of characters—known only to the MNO and not to Cooper/Smith—and then running it through a one-way hashing function. This generates a value for each phone number that is unique for all practical purposes, while making it mathematically impossible to reconstruct the original number. A person with access to the data and the phone number of a subscriber would not be able to find that person in the call records.
[5] The two MNOs exist as a duopoly, with the one under consideration holding more than half of market share.
[6] Contracts for Collaboration website: https://contractsfordatacollaboration.org/

Sustainability and Support to the Government

Cooper/Smith approached this project inclusively, closely involving MoH stakeholders to ensure that outputs of the project were adopted and utilized. As of August 2020, users from the district level to the national level of the Malawi health system were accessing and utilizing the dashboards built by Cooper/Smith to monitor the COVID-19 epidemic.

To ensure sustainability and ownership, Cooper/Smith created a capacity-building plan for the MoH launched in January 2021, which included systemic, organizational, and individual components. As of April 2021, MoH staff has received training to ensure effective knowledge transfer of data processing infrastructure maintenance, updates, and ongoing support requirements. As non-health-related use cases emerge (disaster management, agriculture, education), the government of Malawi (GoM) will need to develop analytical capabilities outside of the MoH to serve different stakeholders. Cooper/Smith, along with the GoM, will explore options to develop capacity within an agency that will ideally be placed to serve all branches of government (e.g., the National Statistical Office (NSO) or the National Planning Commission (NPC)).

Lessons Learned and Recommendations

The outcomes of this project showed that MNO data improves decision-making for public good. Usage by the MoH of Malawi demonstrated that the mass gathering event dashboard was a useful tool for identifying potential hotspots of COVID-19 transmission. Additionally, the epidemiological model, which was enhanced through the use of mobile phone data, proved to be a critical tool in understanding the historical and future trajectory of the COVID-19 burden.

However, the reach of these tools was limited by the lack of additional country-specific COVID-19 data, such as clinical characteristics, risk factors for severity, hospital and ICU use, and all-cause mortality data. Having information on these data points, especially at the district or sub-district level, would have provided valuable additional markers for calibrating and validating the models.

Having information on these data points, especially at the district or sub-district level, would have provided valuable additional markers for calibrating and validating the models.

Next Steps

Cooper/Smith, the MNO, and key MoH stakeholders will continue to support sustainability requests, such as ongoing MoH CDR data access, safe and reliable in-country data hosting solutions, and additional use cases outside of health. In addition, Cooper/Smith and DIAL will closely collaborate on the design of a potential sustainability model to institutionalize this and future data investments across the GoM.

Further possible steps include engaging government stakeholders to explore new, non-health-related use cases and define how to integrate the use of such analyses into existing processes. Cooper/Smith and DIAL are also evaluating the design of an analytical infrastructure that would use novel data (e.g., CDRs) to provide different branches of government with analysis and modeling products. This structure could be located within an agency based in Malawi (e.g., NSO or NPC).

BACKGROUND

In 2017, DIAL, Cooper/Smith, and Infosys embarked on a collaborative effort with the Malawi Ministry of Health and the Malawi Communications Regulatory Authority (MACRA) to demonstrate how mobile network operator data can be used for public good. The working hypothesis was that MNO data and analytics provide a novel and insightful stream of routine information that could be triangulated with more traditional datasets to enhance policy; decision-making; and, ultimately, access to core public services. Up to this point, leveraging MNO data for broader social benefit was still in its infancy.[7] The goal of this project was to develop and test a proof of concept in Malawi.

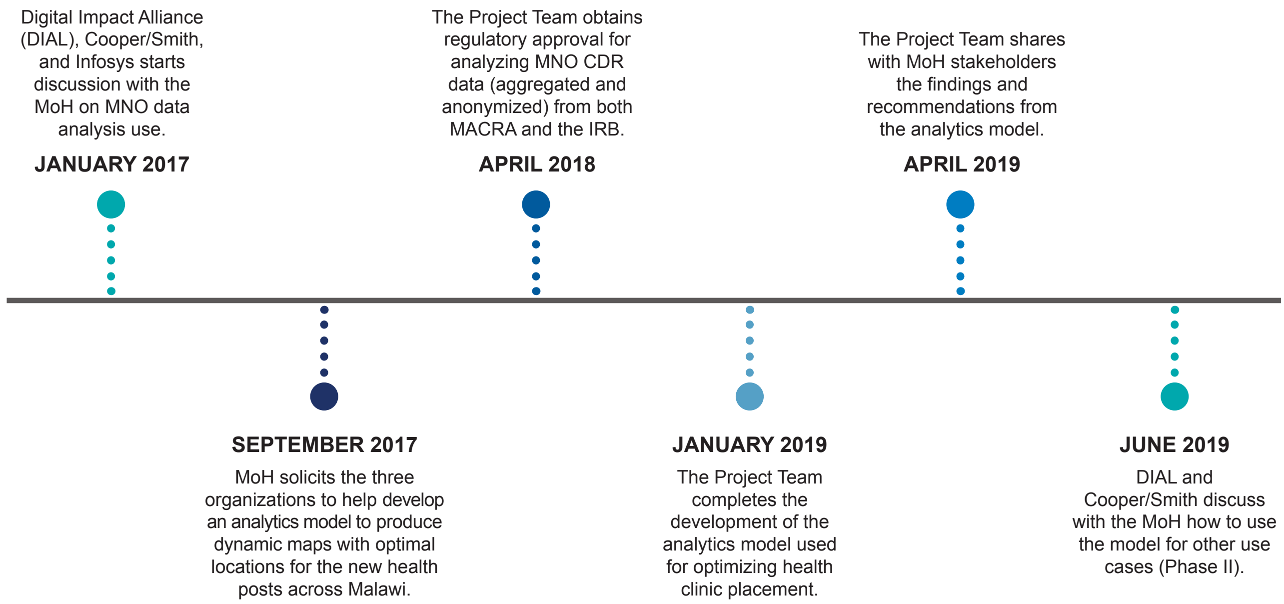
Many stakeholders across the health and development sectors were consulted during project planning to identify pressing challenges and information gaps where MNO data might add value to decision support. A series of use cases were described and evaluated to select the most appropriate for the proof of concept. The selected primary use case aimed to assist the MoH in optimizing its placement of new health services and facilities.

[7] Data-Pop Alliance Primers Series, “Big Data & Development: An Overview”, Emmanuel Letouzé, March 2015: https://datapopalliance.org/wp-content/uploads/2015/12/Big-Data-Dev-Overview.pdf

Phase I

FIGURE 1

Project Phase I Timeline (Jan 2017 - June 2019)



Equalizing Access to Health Services

In 2017, the MoH launched a Capital Investment Plan that included a proposal to build 900 new health posts across the country between 2020 and 2023. At the time, half of the population in Malawi lived more than five kilometers from a health facility, severely limiting access to essential health care services. One primary objective of the CIP was to strategically identify locations for new health posts, thereby ensuring that at least 95% of the population would be living within five to six kilometers of a health facility by 2023.

To achieve that goal, the MoH solicited input from Cooper/Smith to develop a more sophisticated model for optimizing the placement of new health posts that would be constructed in phases as funding became available.

Novel Data Sources and Advanced Analytics

The model used a combination of previously untapped analytics from MNO data and other datasets to understand population growth, density, and migration in Malawi and to optimize the placement of new health facilities in line with Malawi's national priorities. To develop this model, the three organizations obtained and analyzed several datasets, including:

- **Anonymized MNO data provided by a telecom partner**, including CDRs and geotagged locations of cell towers
- **High-resolution population density data for 2015** compiled by WorldPop[8] and calculated using satellite imagery trained on Malawi's previous census data from 2009
- **Location and catchment area calculations of existing health facilities**, provided by the MoH from a detailed UNICEF survey of operating health facilities in Malawi
- **Monthly facility-level disease burden data** provided by the MoH from the national Health Management Information System (HMIS)

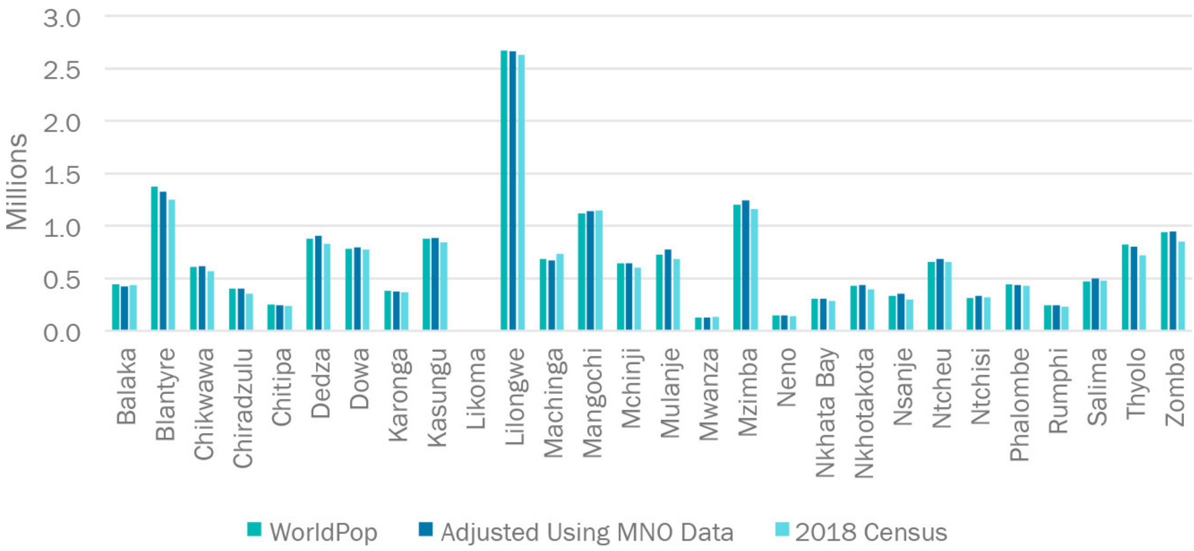
[8] WorldPop Age and Structures/Individual Countries: https://www.worldpop.org/doi/10.5258/SOTON/WP00538

To validate this methodology, the predicted population levels were compared with the results from the 2018 census. The census provided a detailed headcount of population based on on-the-ground surveys at the national, district, and traditional authority levels. This was compared with the population levels based on:

- Static WorldPop data projected uniformly forward for 2016 and 2017[9]
- TA-specific growth rates inferred from shifts in the number of unique users

FIGURE 2

Validation of 2018 Population Projection*



**No data available for Likoma TA*

The results confirm that unique user IDs offer a good proxy for population levels, with a less than 5% discrepancy between the census and the projected population using MNO data.[10]

Key Findings

The model provided the MoH with the following information:

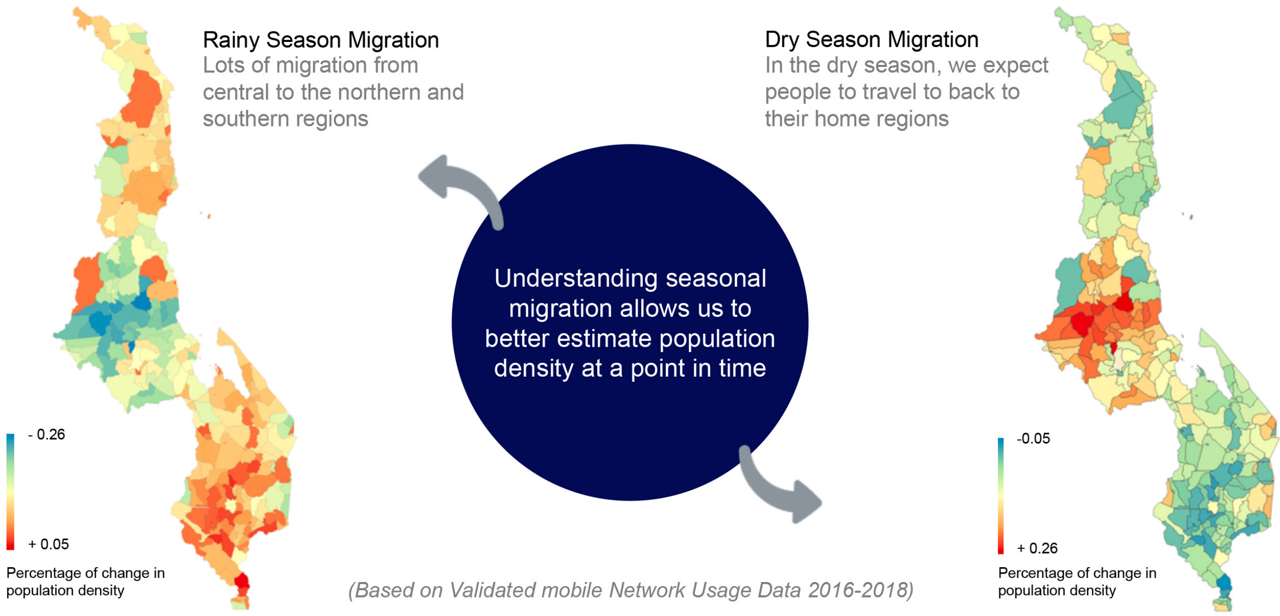
- An estimated 7.74 million Malawians, or 44.7%, live more than five kilometers walking distance from a clinic or health post
- Observations from the deidentified MNO data analytics suggest there is significant population movement on weekends and during the rainy season (*Figure 3*), potentially impacting where health services should be offered and at what time
- The number of people serviced per new health post—a measure of allocation efficiency—varies significantly across districts
- By optimizing clinic placement using the new model, Malawi could achieve 95% coverage targets by 2023 if all 900 posts are constructed (*Figure 4*)

[9] Based on the year-by-year projections WorldPop made available, Cooper/Smith inferred that these projections assumed a uniform annual growth rate of 2.785% across all administrative areas.

[10] While this validates long-term growth patterns, as Cooper/Smith does not know the composition of cellphone users versus those who tend to migrate, short- and medium-term population mobility may be overstated or understated.

FIGURE 3

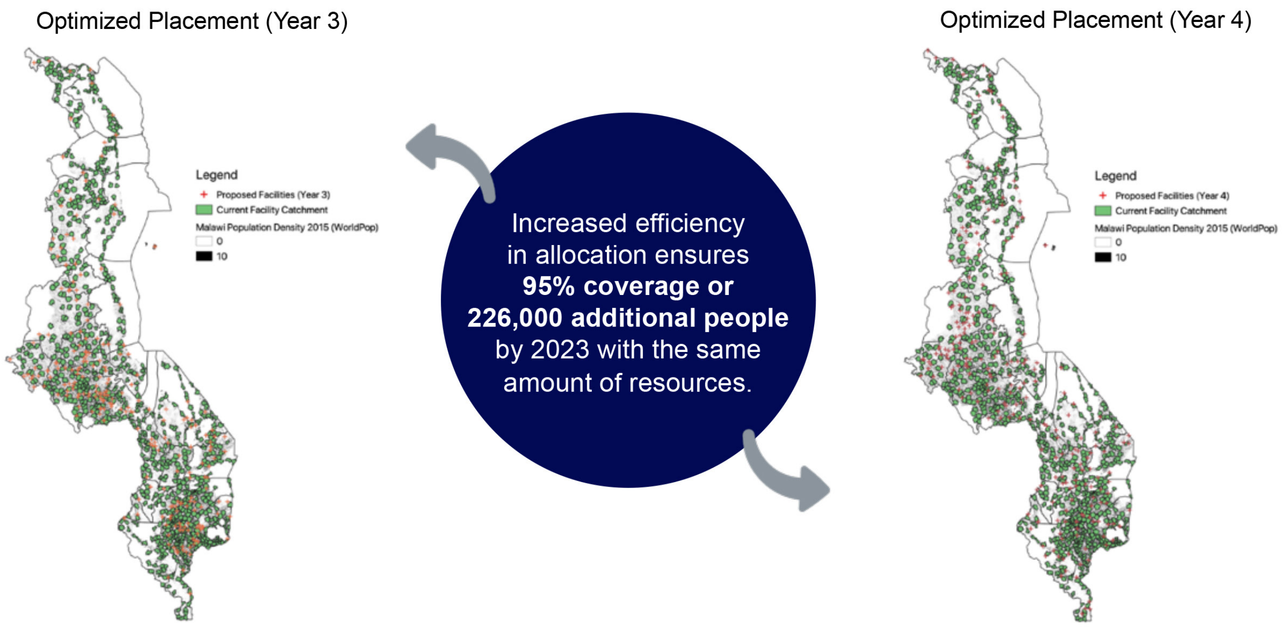
Population Seasonal Mobility Patterns



(Based on Validated mobile Network Usage Data 2016-2018)

FIGURE 4

Proposed Allocation of New Health Posts



Country Use and Sustainability

At the conclusion of Phase I, the MoH was still mobilizing resources to construct new facilities as prescribed in the CIP. The Department of Planning and Policy Development (DPPD) indicated a strong desire to utilize the initial findings to locate clinic placement and to receive routinely updated results in an easy-to-digest format. DPPD wished to consult the model and updated results as new funding became available for the CIP. The positive reception and further requests for functionality were considered in the planning and execution of Phase II, described below.

Phase II

In April 2020, Malawi confirmed its first three cases of COVID-19 (*see COVID-19 epidemic in Malawi timeline in Figure 12, page 24*). The MoH needed to quickly leverage existing data sources and improve models to respond to the pandemic in a data-driven way. At the request of the MoH, Cooper/Smith pivoted focus and resources to accelerate the country's COVID-19 response.

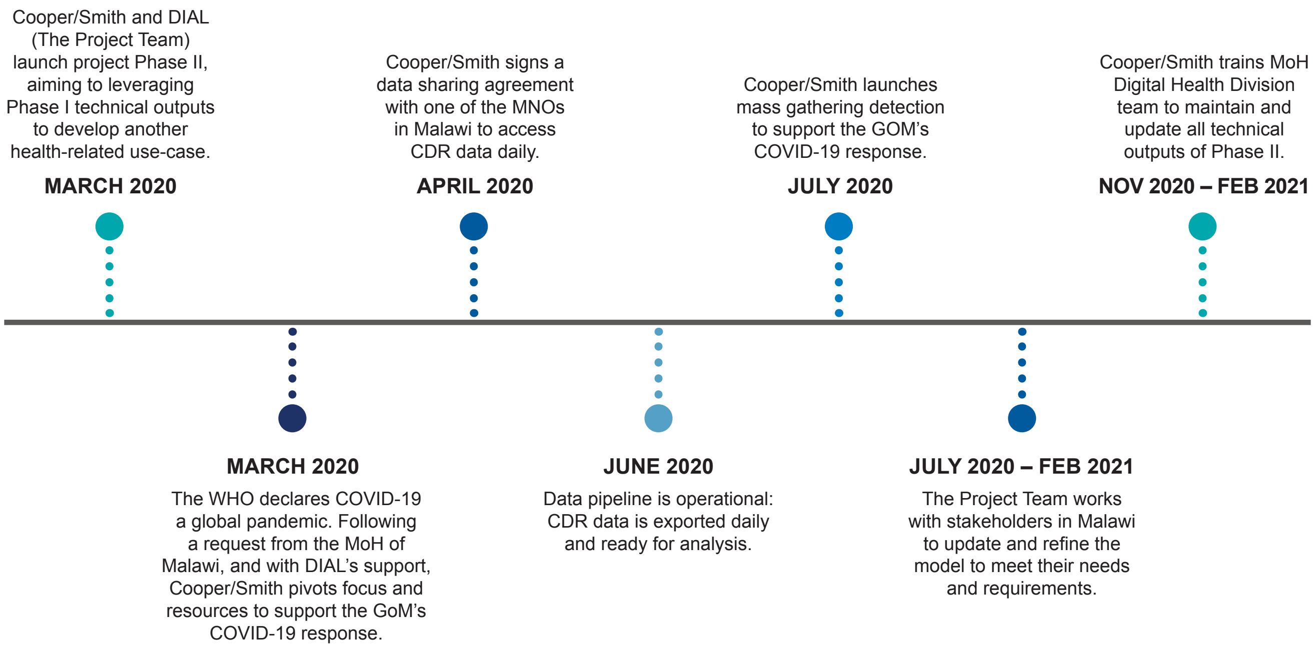
Building on evidence from Phase I that MNO data can be a useful proxy for population density and movement, it was clear that these data could be a valuable asset in mounting a rapid and targeted public health response to the pandemic. Cooper/Smith and the MoH quickly leveraged existing partnerships with MNOs, accessing additional CDR data that could inform response efforts.[11] The resulting work led to key population insights in near-real time to support government policy and decision-making during this critical period.

This report describes:

- The regulatory, technical, and legal frameworks set up by Cooper/Smith to ensure privacy and accountability for using MNO data during the COVID response
- The methodology and models used to generate key insights
- The use cases and associated technical products developed to support the MoH with decision-making

FIGURE 5

Project Phase II Timeline (March 2020 - February 2021)



[11] The data is taken from March 2020 to April 2021

REGULATORY, LEGAL, AND TECHNICAL FRAMEWORKS

The novelty of working with data and computational sciences in the development field, coupled with increasing surveillance concerns, has heightened the need for stakeholders to prioritize compliance with best practices in data management to mitigate concerns over privacy and ethics. Protection of privacy and strict avoidance of unintended uses of data are essential to ensure trust and proper use as well as protect individual and group privacy.

From Phase I inception in 2018, Cooper/Smith followed a rigorous approach to ensure that all efforts and technical work abide by local regulations and apply best practices to protect data privacy, ensure responsible data use, and follow best practices regarding ethics in data science.

Regulatory framework

In early 2018, Cooper/Smith obtained authorization from the Institutional Review Board (IRB) of the National Commission for Science & Technology of Malawi to perform a mobile data analysis in the country. The NCST advises the GoM and other institutional actors on all matters related to science and technology. Obtaining its authorization ensured the buy-in of all key stakeholders in Malawi. Additionally, the Malawi Communications Regulatory Authority, which enforces data privacy regulations in the country, authorized the MNO to share anonymized CDR data for analysis. With these authorizations, the IRB and MACRA gave Cooper/Smith their approval to proceed with the analyses for Project Phase I and II.

Legal Framework

Cooper/Smith and an MNO[12] based in Malawi signed a data sharing agreement that established a framework and mechanism for how data is shared between project partners. The DSA defined data and anonymization processes, the data sharing period, analysis goals and deliverables, regulations and laws, and more. The DSA also specified that Cooper/Smith accessed and anonymized data for its analysis and shared the outputs of that analysis, and that no data could be shared with any third party. Contracts for Collaboration provided examples of DSA agreements between MNOs and third-party organizations.[13] This agreement made it possible for Cooper/Smith to receive CDR data from the MNO daily.

Cooper/Smith and an MNO based in Malawi signed a data sharing agreement that established a framework and mechanism for how data is shared between project partners.

Technical Framework

Prior to accessing any data, Cooper/Smith developed an MNO data processing guide that provided additional technical details on the type of data needed; the anonymization processes; the data journey from MNO raw data to aggregated and anonymized data from Cooper/Smith using a data pipeline; and Cooper/Smith's hosting infrastructure and security measures that were put in place.

The guide was reviewed by both the MNO and the UNOCHA team, which reviewed and assessed the project data privacy and ethics compliance. Both organizations deemed the infrastructure adequate and secure.

---

[12] The two MNOs exist as a duopoly, with the one under consideration holding more than half of market share.
[13] Contracts for Collaboration website: https://contractsfordatacollaboration.org/

ENSURING THE RESPONSIBLE AND ETHICAL USE OF DATA

Projects that leverage novel data sets, especially from sensitive sources like mobile phones, demand an elevated ethical and legal approach that goes above and beyond existing laws and regulations to ensure the protection of human rights and privacy. All actors involved in this project took extensive care to ensure the respect and enforcement of data privacy regulations and avoid any misuse of the data and the model outputs.

The data used in this project was fully anonymized by the MNO before being shared with Cooper/Smith. In addition, Cooper/Smith built a highly secure data pipeline that went beyond industry standards. The data was stored in a secure environment that did not permit the transmission of data at the individual level.

Following the Principles for Digital Development,[14] Cooper/Smith worked with other stakeholders at every stage of project development to ensure that ethical standards were met and that the limitations of the model were well understood. With the mass event model, it was made clear that predictions were meant as an additional source of knowledge rather than an authoritative guide to policy response. It would have been both inappropriate and counterproductive, for example, to order the lockdown of an area based on a single report from the model.

The user interface for the data dashboards was also designed with these ethical considerations in mind. The layout and visualizations were chosen to give the user a quick overview of a flagged event, the information necessary to situate it in context, and the ability to quickly decide which events merited follow-up or response.

[14] Principles for Digital Development: https://digitalprinciples.org/

Security

Cooper/Smith only received data after it had been fully anonymized and did not have access to phone numbers or timestamps at any stage of the process. As outlined in the sections on CDR data and anonymization (*page 18*), ancillary data (*page 19*), and security infrastructure (*page 19*), Cooper/Smith set up a robust security infrastructure to ensure individual data was never accessed. Data was also structured to minimize the risk to the subscriber even in the case of a data breach.

Data Completeness and Bias

Models are imperfect representations of reality because of both the assumptions of their authors and the biases and limitations in the data they use. Research by the Global System for Mobile Communications Association (GSMA) has shown that in sub-Saharan Africa there is a 15% gender gap in mobile ownership (GSMA, 2019). In the case of mobile data in Malawi, women, young people, and the elderly may have been underrepresented in the model since adult men are more likely to own a phone. This model based on mobile phone subscribers' mobility also likely overrepresented the movements of men, who are more likely to own a mobile phone, compared to women, who are more likely to visit health facilities but less likely to own a mobile phone (GSMA, 2019). Additionally, since the data only records calls, those with higher incomes were likely to make more calls and thus would appear in the data set more frequently.

Aware of these limitations, Cooper/Smith limited the model's use to relative movement. A mass event was defined in comparison to previous data for that location (meaning a difference in the number of subscribers located in the same cluster between two different time periods) and the interchange scored in the microsimulation was only relative values between two clusters (meaning the number of subscribers moving from one cluster to another one).

The mobile data was used to seed the SEIR model (meaning to decide when an infection would start in a specific cluster), but the model itself took age groups into account based on census data and calculated infection independent of the mobile data. This ensured that the mobile data's biases were not inherited by the SEIR model.

In the case of mobile data in Malawi, women, young people, and the elderly may have been underrepresented in the model since adult men are more likely to own a phone. Aware of these limitations, Cooper/Smith limited the model's use to relative movement.

Risk/Benefit Self-Assessment

Cooper/Smith completed a risk/benefit self-assessment on the population mobility model in December 2020. This process is part of a broader DIAL risk and responsible data management strategy to ensure that data is being used responsibly within DIAL-supported projects. It is designed to help teams identify and document risks to their project and its subjects and how to avoid, mitigate, or manage them.

The assessment was conducted through a qualitative review of project documentation, identification of concerns and questions, and discussion with the Cooper/Smith team. The panel identified risks to vulnerable populations based on their underrepresentation in the data. It also pointed out possible ways underrepresentation might modify overall trends and reduce the model accuracy. It recommended investment in monitoring and evaluation in order to assess intended and unintended outcomes from the project and engagement and adoption by key stakeholders in Malawi. Taking into account the mitigation measures that were implemented by Cooper/Smith, the risks associated with unauthorized data disclosure, legal issues, and data quality for the project were assessed as relatively low.

External Model Assessment

UNOCHA has created a peer review framework for predictive analytics in humanitarian response that "aims to create standards and processes for the use of models in our sector. It is based on research with experts and stakeholders across a range of organizations that design and use predictive models" (UNOCHA, 2020).[15]

Between November 2020 and April 2021, UNOCHA also conducted an independent review of the project, assessing the model, how it was implemented, and whether ethics considerations and potential risks were taken into consideration and mitigated. The assessment concluded that the mitigation plans set up by Cooper/Smith were sufficiently robust to mitigate the identified concerns.

Ethics Panel on Novel Data Use in Malawi

Government ministries and others in the data for development field recognize the vast potential for novel data sources to address key challenges, along with the critical need for ethical protections to be initiated from the outset. In addition to the above assessments conducted through this project, Cooper/Smith and Georgetown University, with support from the Gates Foundation, are working with low- and middle-income countries (LMICs) to establish the Ethics Panel Project, which taps into these deep repositories of mobile data responsibly, while minimizing risk and unintended consequences in maximizing privacy protections, transparency, and accountability.

The Ethics Panel Project is working to convene experts to help develop an evaluation framework for the use of novel data in public health programs, using lessons learned from this project and others. Thematic areas will include consent, data storage and access, data transfer, identification and reidentification, analytic methods and indicators, program application, and communication.

The Ethics Panel Project is also expected to support and advise on the use of mobile phone data by ensuring best practices in research and implementation are adhered to when collecting and analyzing sensitive information. The methods and approach developed under this project will be shared with the panel for review and potential use as recommendations to the GoM.

Government ministries and others in the data for development field recognize the vast potential for novel data sources to address key challenges, along with the critical need for ethical protections to be initiated from the outset.

[15] The full report is available at https://centre.humdata.org/catalogue-for-predictive-models-in-the-humanitarian-sector/

METHODOLOGY AND RESULTS

Empirical COVID testing and case management data in Malawi suggested differential timing for when the epidemic would begin in each district, village, and even neighborhood. Although many hypotheses existed as to why the cases would appear in one location over another, the consensus was that rate of intermingling between local populations was likely the core driver.

To test this hypothesis, Cooper/Smith conducted an exploratory data exercise, building on the methods from Phase 1 (DIAL, 2020). Two promising use cases were identified:

- **Simulate the spread of the pandemic across the cellular network to improve epidemiological estimations**
- **Detect anomalous mass gatherings to identify potential transmission events and tailor public health response locally**

To develop these use cases, Cooper/Smith built a secure and efficient data pipeline and processing infrastructure, detailed below.

Data Pipeline

The main data source analyzed was a list of CDRs provided by the MNO. Each of these represents a set of outgoing voice calls or SMS messages sent from a particular tower.

CDR Data and Anonymization

Before the MNO shared data with Cooper/Smith, the data was preprocessed to ensure privacy protection and quality assurance. This was executed in two primary ways: First, all identifying data collected by the MNO, such as name, phone number, gender, and age, was stripped out and each user was assigned an individual code unrelated to their phone number.

Approximately 1.5 million call detail records are received daily, amounting to 5.6 billion records received as of April, 2021.

Anonymization of phone numbers was done using an MD5 hashing algorithm and a salt value. This involved appending an arbitrary string of characters—known only to the MNO and not to Cooper/Smith—and then running it through a one-way hashing function. This generated a value for each phone number that was unique for all practical purposes, while making it mathematically impossible to reconstruct the original number. A person with access to the data and the phone number of a subscriber would not have been able to find that person in the call records.

Secondly, the timestamp was removed and replaced with a four-hour block, which prevented potential triangulation of a user's location and movements. If exact timestamps were used, then a physical observer could have seen exactly when a user used their phone and could guess which cell tower they might have been near. The observer could then find a user ID in the data that matches that behavior. With specificity reduced to a four-hour window, many more people would match a particular pattern, making triangulation impractical.

With these protections in place, identifying information was only seen by MNO employees and Cooper/Smith, which received a CDR that included subscriber ID, tower code, date, time, and number of calls:

FIGURE 6

Example of CDR Structure Received by Cooper/Smith

Subscriber ID	Tower Code	Date	Time	# of Calls
abcd1234ef	CELL1234	Jan 2, 2021	12:00-16:00	4

These data, approximately 1.5 million records, are received daily. Since March 1, 2020, as of the time of writing (April 2021) this amounts to 5.6 billion records.

Ancillary Data

Several other data sources were used to inform the models:

- The GPS coordinates of all the cell towers provided by the MNO
- A country map of the population density estimates provided by WorldPop[16], a project of the University of Southampton
- A shapefile of Malawi's administrative divisions and UN estimates of population growth since the last census, which provides a grid of 1km² tiles and estimates how many people live in each tile, with the results provided as a TIFF raster file

Security Infrastructure

Cooper/Smith abided by all local regulations and applied best practices to protect data privacy, ensure responsible data use, and follow best practices regarding ethics in data science. Overall, Cooper/Smith followed two sets of best practices regarding:

- Web application[17] and cloud security:[18] data encryption; password management; controls and procedures; etc.
- The General Data Protection Regulation (GDPR)[19] regarding personal data anonymization and data protection by design

To protect and respect Malawian data, Cooper/Smith created a robust security infrastructure to host the data. Personal data was removed, and the records were anonymized on MNO servers in Malawi before being sent to Cooper/Smith. A virtual private cloud (VPC) was set up with no direct exposure to the external internet, and a bastion server[20] was its only entry point.

A service on this VPC monitored a drive at the MNO, and when the daily file appeared, it was downloaded onto the VPC. There, it could only be accessed by administrators via the VPC—all other connections were blocked.

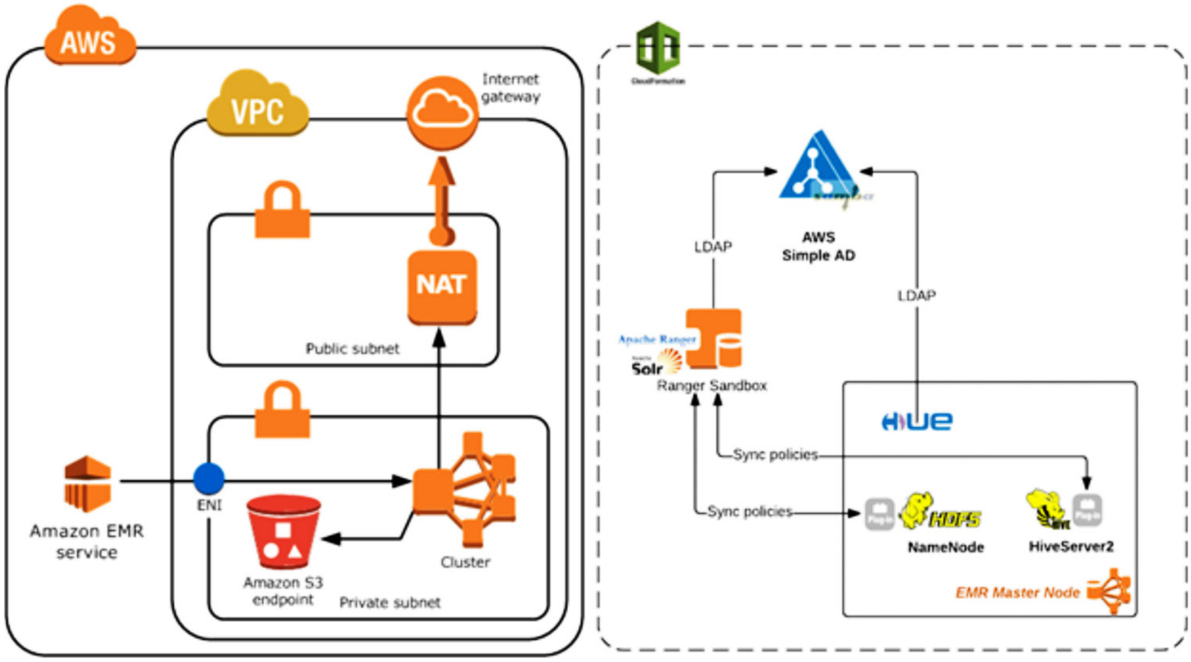
---

[16] WorldPop Age and Structures/Individual Countries: https://www.worldpop.org/doi/10.5258/SOTON/WP00538
[17] Web Application Security Best Practices: A Developer's Guide: https://securityintelligence.com/posts/web-application-security-best-practices-developers-guide/
[18] 21 Best Practices for AWS Cloud Security: https://www.cypressdatadefense.com/blog/aws-cloud-security-best-practices/
[19] European Commission Data Protection Law: Personal data (https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en) and data protection (https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/what-does-data-protection-design-and-default-mean_en)
[20] Within a network, a bastion server is specifically designed and configured to withstand attacks. A bastion server is an instance used as the public entry point for accessing a private network from the internet. Traffic flows through the bastion server to access the private network. The bastion server has security mechanisms to manage that traffic.

As shown in Figure 7, the data was stored in an Apache Spark database. Due to the sheer volume of records (1.5 million received daily), processing was done on an eight-core cluster of servers, also entirely inside the VPC. This produced aggregate statistics at the cell tower level, which was the only data allowed to leave the VPC. Individual-level data was not permitted to leave. This was accomplished by having not only table-specific but also column-specific access rules on the database, where users with remote access, such as dashboard software, were only allowed to see aggregated data.

FIGURE 7

Data Processing Infrastructure



Dashboard visualizations were provided by Tableau Online as a platform that did not allow for user access of personal data sources, but rather displayed the final outputs of the analysis. As such, Cooper/Smith used a virtual Microsoft Windows server and the official Tableau bridge client to push the aggregated data from the VPC to Tableau Online.

To ensure security, this methodology supported outgoing connections only, not incoming ones. To make the connection secure when connecting to Tableau Online, Cooper/Smith blocked incoming traffic from internet or other network segments using inbound firewall rules and opened outgoing connections using outbound firewall rules to connect to Tableau Online.

Data Processing

For the MNO data to be usable by models, Cooper/Smith had to process, clean, and validate that it was representative of the population in Malawi.

This involved several steps, outlined in the sections below:

- Clustering the towers
- Projecting the subscriber population of each cluster onto the administrative areas of Malawi
- Comparing the predicted populations to UN estimates to show correlation

Tower Clustering

In cellular networks, most towers are a collection of antennas mounted in a unique location, with each antenna appearing as a unique tower in the data. As a result, many towers have identical or near-identical GPS coordinates. Cooper/Smith used a simple clustering algorithm to ensure that antennas on the same tower, or extremely close towers on adjacent buildings (within one thousandth of a degree, or approximately 110 meters) were treated as a single location. This algorithm reduced more than 10,000 unique tower IDs to just 729 clusters.

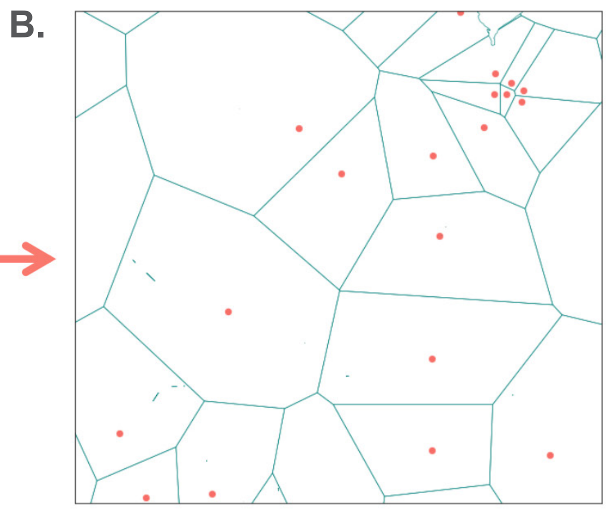
Each cluster is a point on a map representing the centroid of the towers in that cluster.

To compare that with a population map, Cooper/Smith had to calculate a coverage area for each cluster. To do so, it generated a map of Voronoi cells for each cluster, where every point in Malawi is assigned to the nearest cluster centroid. These varied widely in size between urban and rural areas. For example, in Mchinji District, traditional authorities (TA) and cells have an average area of 266km² and 223km² respectively. By contrast, in Lilongwe City, wards have an average area of 7.0km² and cells an area of 5.1km². The result is 729 separate cells as shown in Figures 8-A and 8-B below:

FIGURE 8

A. Malawi Divided Into Its Cellular Network Using the Voronoi Method

B. Detail of Cell Towers and Their Coverage Cells



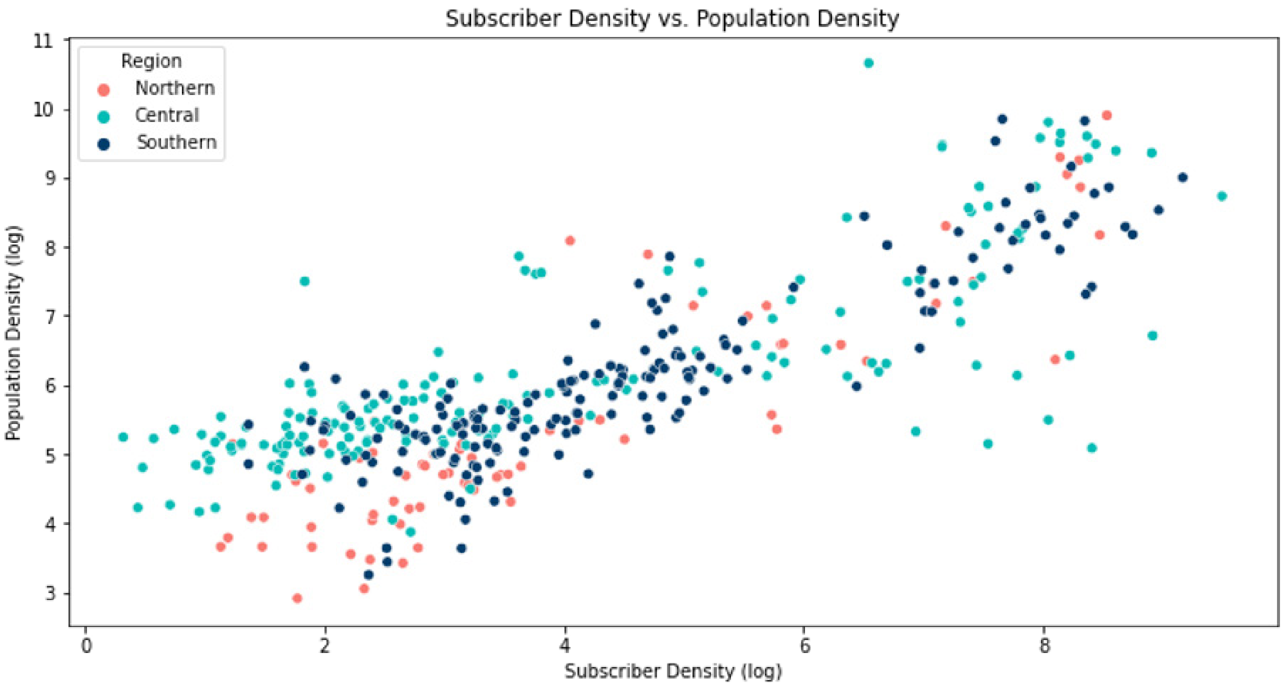
Validating the Model

To use subscribers as a proxy for the population, it was necessary to prove that they were correlated with actual population figures. This was completed between 2016-2018 in Phase I of the project (DIAL et al., page 15-16), when data was leveraged from a different MNO that provided a different set of data. As such, the correlation to actual population figures had to be redone and refined for Phase II.

Cooper/Smith did this by calculating what the predicted subscriber population of each TA would be by distributing the cluster subscribers to each TA by share of overlapping area. Cooper/Smith then compared the logs of these predictions and the predicted UN populations for 2020, giving a correlation with an R^2 of 0.67 and a mean squared error of 0.83. Bodies of water, nature reserves, and national parks were removed from the data for these purposes, as seen in Figure 9 below:

FIGURE 9

Comparison of Population Density and Subscriber Density as Calculated by Using Overlapping Area



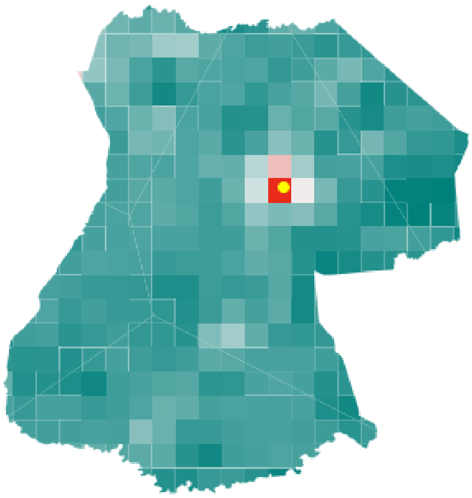
With this, Cooper/Smith confirmed that the subscriber population correlated with the actual population. Predictions were then improved by using the density maps provided as TIFF raster files by WorldPop,[21] which provided the estimated population in each 1km² tile of the country.

In rural areas, for example, a cell tower is likely to be in a major population center, having a high density close to the tower and a much lower density in the rest of the cluster's coverage area. (*See Figure 10.*)

[21] World Pop Population Counts, Individual Countries, Malawi 100m. population: https://www.worldpop.org/doi/10.5258/SOTON/WP00538

FIGURE 10

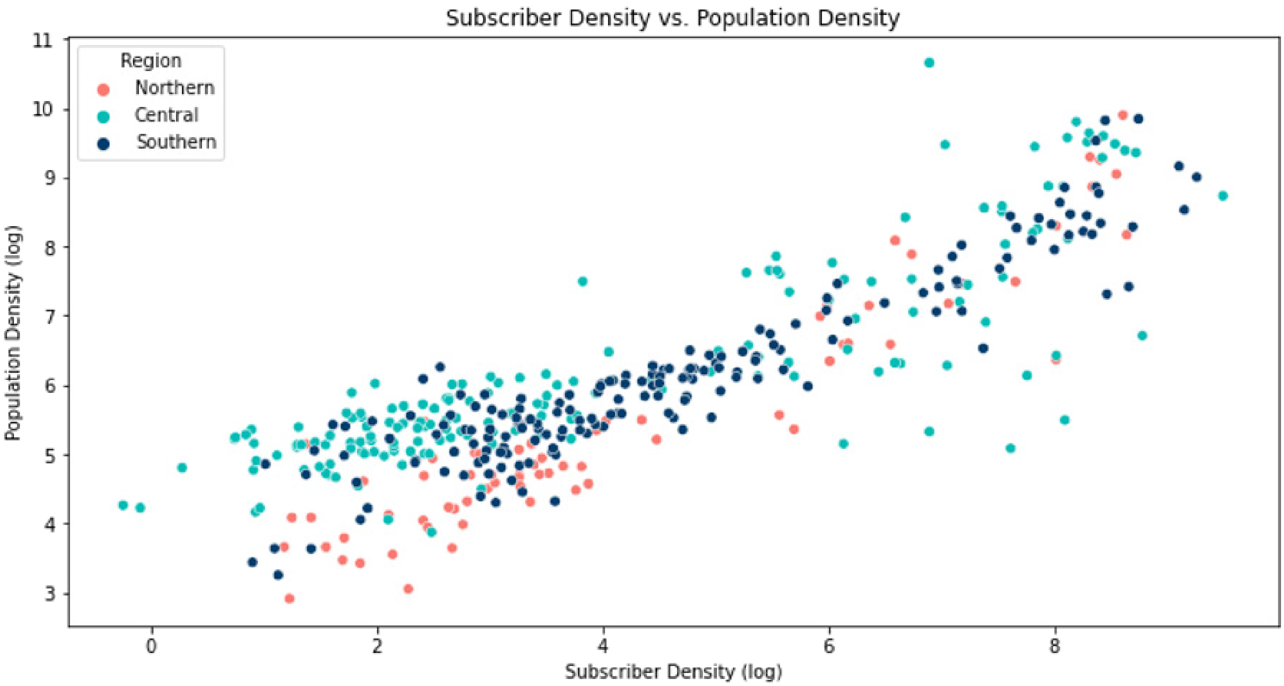
Example of a Rural TA Whose Only Major Center of Population Is Also the Location of the Cell Tower



Using this raster map, Cooper/Smith distributed a cell's subscriber population using a density map instead of overlapping area. This resulted in an improved R^2 of 0.75 and a mean squared error of 0.71, as shown in Figure 11 below:

FIGURE 11

Comparison of Population Density and Subscriber Density as Calculated Using a Density Map



This process allowed data to be displayed by TA rather than by cell cluster, facilitating its use by decision-makers and allowing the data to serve as an input to other models that use TAs as their focal geography, such as the epidemiology model described on page 24.

Use Case #1: Enhancing Epidemiology Estimates

Background

In mid-March 2020, Cooper/Smith developed a mathematical model of COVID-19 transmission like those created by other institutions, such as Imperial College London, University of Washington's Institute for Health Metrics and Evaluation, and other government entities and research institutions (American Hospital Association, 2020; Reiner et al., 2021; Walker et al., 2020).

While Malawi had not yet detected its first case, it had declared a national state of emergency under the presumption that COVID-19 spreading to Malawi was imminent. (*See Figures 12 and 13 for a timeline of the COVID-19 epidemic in Malawi and the resulting cases and deaths.*) Although the government had begun emergency operations support, it had not undertaken an effort to simulate what a COVID-19 outbreak could look like in terms of infections, health system burden, and deaths.

FIGURE 12

COVID-19 in Malawi (March 2020 - February 2021)

MARCH 2020
The government declares COVID-19 a disaster. It closes public and private schools, cancel public events, and ban gatherings of more than 100 people.

APRIL 2020
First three COVID-19 cases are confirmed in Malawi.

APRIL 2020
The government announces a 21-day natioinal lockdown to be put in place, but under pressure from the population, decides to cancel it.

MAY 2020
The government tells all people except essential workers to stay at home.

JUNE 2020
President Lazarus Chawera enters office.

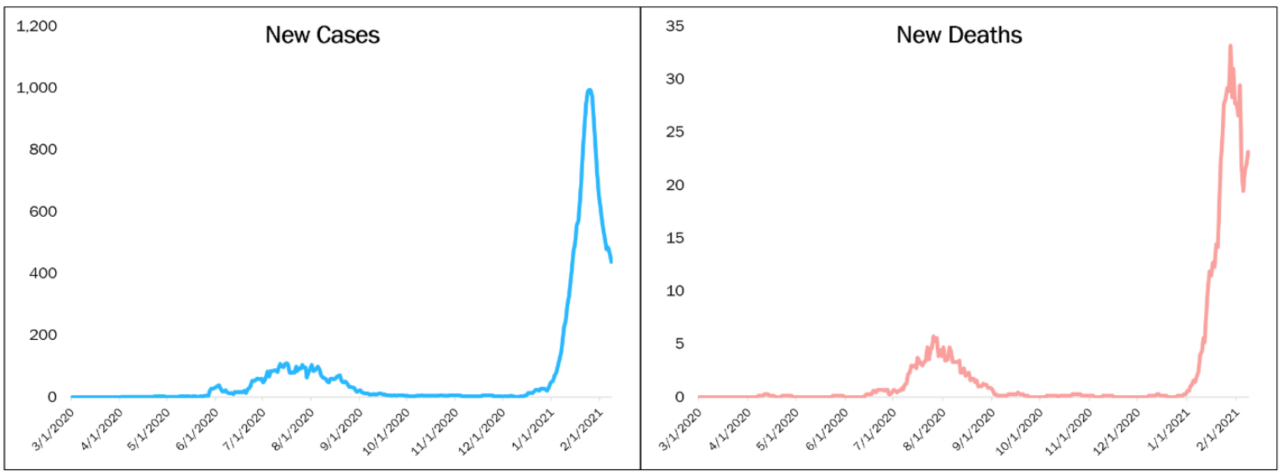
JULY 2020
President Chakwera reconstitutes the Presidential COVID-19 Taskforce and sets up the National COVID-19 Office. This office introduces new measures such as mask mandates and social distancing.

JANUARY 2021
President Chakwera declares a state of national disaster following a second wave of cases, imposing school closures, a nighttime curfew, and a ban on gathering of more than 50 people.

FEBRUARY 2021
As of February 12, 2021, Malawi has recorded about 28,628 confirmed cases and 935 deaths.

FIGURE 13

Seven-Day Moving Average of New COVID-19 Cases and Deaths (March 2020 - February 2021)



With guidance from the Digital Health Directorate and the Public Health Institute of Malawi, Cooper/Smith began developing an open-source model to simulate an outbreak in Malawi down to the TA level for use in emergency planning and response (Cooper/Smith, 2020).[22] The initial purpose was to anticipate trends in infection, mortality, and health system use. As the first infections were detected in Malawi in early April, the modeling priority shifted to understanding the potential effects of non-pharmaceutical interventions such as physical distancing, more extreme lockdowns, school and business closures, and use of cloth masks.

Epidemiological Microsimulation

Without empirical data on "patient zero" for each TA, Cooper/Smith had to estimate the epidemic start date, first by leveraging estimates of interchange, or connectedness, between TAs from the CDR data. These interchange scores can be interpreted as the fraction of mobile subscribers in one TA who visit another TA. Cooper/Smith then seeded a microsimulation with known COVID diagnoses and ran a simulation where each TA with an infectious population exerted infection pressure on all the others according to their interchange scores.

This microsimulation generated an expected number of infections for each TA on each day of the pandemic. From this, Cooper/Smith could extract which date it is believed the first (or nth) infection of each TA was to have taken place. Because a single infection may not be sufficient to spur community transmission of COVID-19, Cooper/Smith initially assumed that community spread would occur upon accumulating 10 COVID-19 infections.

Use Case #2: Mass Event Detection

Mass event detection was explored as a second use case deemed meaningful to Malawi decision-makers. Mass events are events that had occurred the previous day and which had abnormally high numbers of MNO subscribers gathering in one place. If these events were verified, the Emergency Operation Center (EOC) could prepare for or mitigate risk by mobilizing health workers to respond in the case of a potential outbreak, quickly getting ahead of the risk of exponential infection.

Mass events are events that had occurred the previous day and which had abnormally high numbers of MNO subscribers gathering in one place.

The MNO data provided users' locations and movements throughout the day, allowing the EOC to determine not only whether the event was a risk for mass infection, but also if the risk had likely spread. This made it possible to generate a "risk score" for areas surrounding the event, distinguishing between events that were mostly regional and those that attracted subscribers from other parts of the country.

Cooper/Smith developed all dashboards according to a user-centered design (UCD) philosophy, which meets Principles for Digital Development #1, *Design With the User*. UCD allows the user to drive development, leading to more intuitive and useful information products. The resulting dashboard was developed in July 2020, giving MoH users the information needed to determine whether a potential event was worth verifying. Based on feedback from users, the dashboard was designed to display only 25 events, preventing information overload and decision paralysis.

Defining a Mass Event

Mass events were detected by identifying anomalous rises in clusters of subscribers. Since the data were grouped in four-hour blocks, Cooper/Smith created a timeseries for each time block and cluster. Then it calculated the number of standard deviations away from the mean for that cluster/block combination (e.g., the z-score), and defined a mass event as having a z-score of greater than 2.5 and a subscriber count of more than 500.

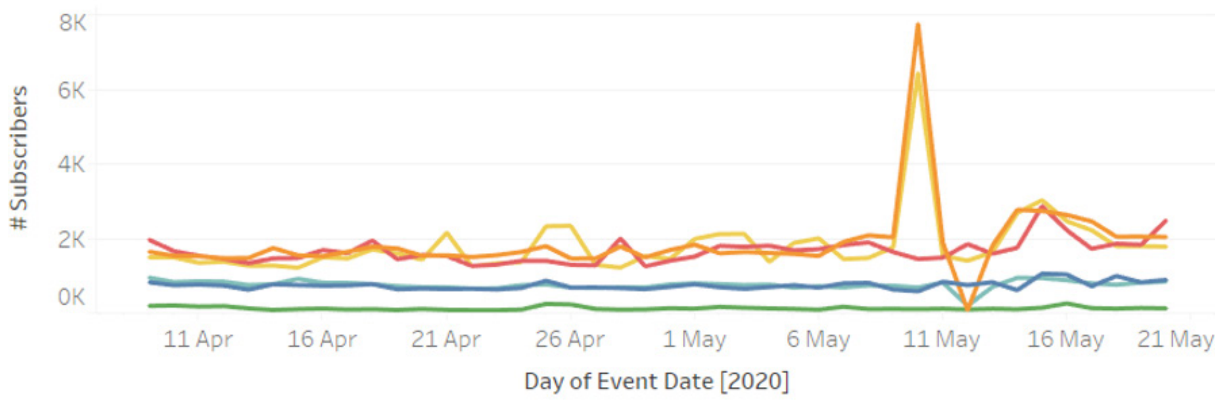
---

[22] Cooper/Smith also developed COVID-19 risk and vulnerability models that are accessible online for free. More information is in Appendix (*page 43*).

This approach was a compromise between competing priorities. Measuring the population of the anomalous event would have resulted in normal fluctuations in dense areas with outsized prominence, while measuring the percentage of the mean would have overemphasized tiny clusters experiencing small movements. When visualized as a time series, users could more easily identify certain kinds of events. For example, as seen in Figure 14, a one-off event that filled a stadium appeared as a large spike in the timeline. This event was confirmed by multiple local news sources.

FIGURE 14

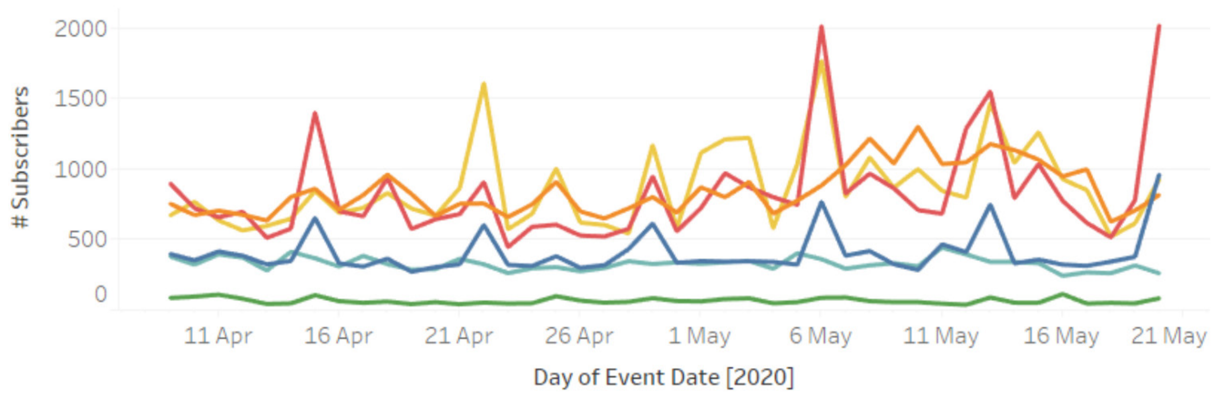
Single Peak in Subscribers for a Large One-Off Event on May 10, 2020, at the Mzuzu Stadium in Malawi



Meanwhile, as seen in Figure 15, recurring events such as markets showed up as regular spikes on a particular day of the week. For example, this market occurred each Thursday:

FIGURE 15

Recurring Peaks in Subscribers for a Weekly Event at the Chimbiya Market in April - May 2020



This classification system was by nature imperfect and relied on human confirmation of each event led by the EOC. False positives could be caused by nearby towers going offline, thereby redirecting traffic to new clusters, or by small increases in nighttime traffic, with the nighttime blocks often showing little fluctuation. However, with a robust human verification system, mass event alerts were a valuable source of information.

Mapping Mass Event Risk

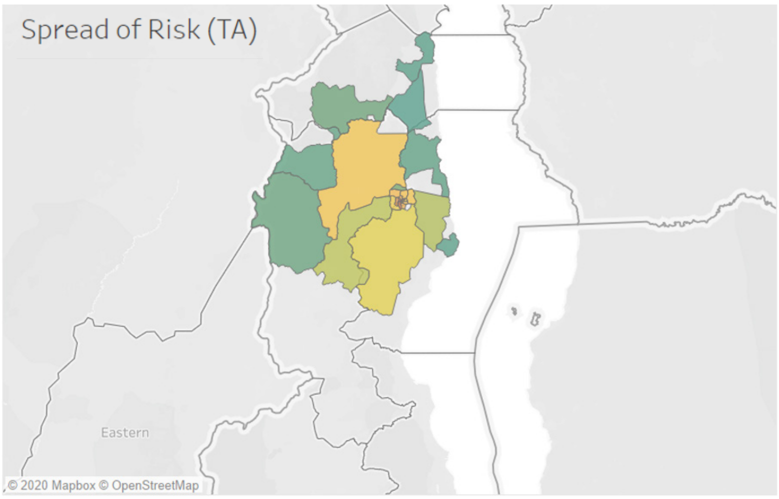
The MNO data contained not only a registry of all the subscribers in a particular cluster at a specific time, but also the other clusters those subscribers visited. Using this data, Cooper/Smith could determine what other areas would be at risk if a mass event was found to be a locus of infection. From there, it was possible to extrapolate to the TA level to be useful to practitioners.

Because the focus was to provide these insights in near real time, Cooper/Smith calculated a cluster's risk score from a mass event as the log of the number of shared subscribers between the two events.

Thus, **if 1,000 people had attended a mass event in Cluster A, and 100 of those were also in Cluster B on the same day, then Cluster B's risk score would be ln (100) = 4.65**. Figure 16 shows an example of an event that attracted mostly subscribers from adjacent TAs.

FIGURE 16

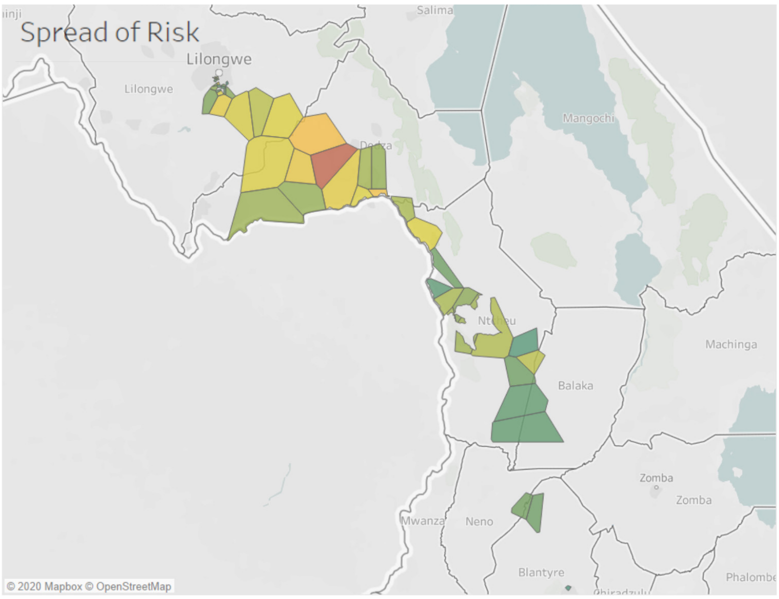
Risk Map of a Localized Event Around Mzuzu, May 2020



By contrast, Figure 17 displays the population movements associated with a market that was visited by subscribers from both Lilongwe and Blantyre, showing the risk stretching much further and following the main roads:

FIGURE 17

Risk Map of an Event With a Wider Risk Profile, Including Both Blantyre and Lilongwe, May 2020



Using the algorithm above, Cooper/Smith mapped the cluster risk scores onto TAs to create risk scores for each TA. Presenting the data this way allowed decision-makers to quickly act on the information, even if the cluster-level information was in some cases more precise.

Dashboard Development

Mass Event Map Overview

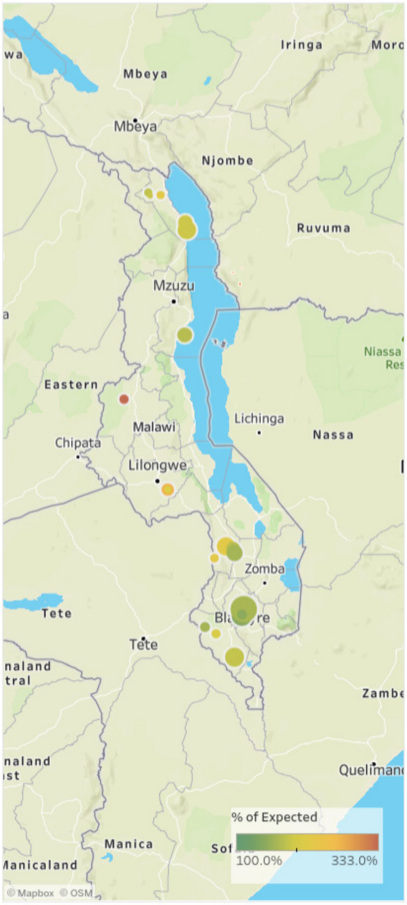
FIGURE 18

Mass Event Overview Map of Malawi, August 2020

Note: Data shown in this image has been scrambled to protect confidential data shared by the MNO

This map shows the most significant mass events that occurred on Friday, August 21, 2020 in Malawi. It shows the top 25 events with a Variance Score of 2.00 or greater.

Click any dot on the map, and the event risk report will be shown below



Active Date
The date of the map and the event report
Friday, August 21, 2020

Active Period
The four-hour block of the event report
08:00:00

Active Cluster
The cluster ID# of the event report
621

Top Mass Events

Daily ..	Cluster Name	Period	Expected Subscribers	Total Subscribers	% of Expected	Variance Score
1	Tower 1	20:00:00	312	761	244.2%	9.45
2	Tower 2	16:00:00	830	1,718	207.1%	6.24
3	Tower 3	08:00:00	753	2,056	273.2%	6.12
4	Tower 4	04:00:00	335	679	203.0%	5.83
5	Tower 5	16:00:00	5,576	9,325	167.2%	5.40
6	Tower 6	08:00:00	223	422	188.9%	4.39
7	Tower 7	12:00:00	760	1,597	210.0%	4.31
8	Tower 8	16:00:00	583	890	152.8%	4.01
9	Tower 9	08:00:00	569	1,130	198.5%	3.96
10	Tower 10	08:00:00	348	1,158	333.0%	3.93
11	Tower 11	16:00:00	3,822	5,423	141.9%	3.89
12	Tower 12	16:00:00	911	1,352	148.4%	3.84
13	Tower 13	08:00:00	2,012	4,431	220.3%	3.83
14	Tower 14	08:00:00	2,609	4,814	184.5%	3.82
15	Tower 15	08:00:00	1,366	2,394	175.3%	3.65
16	Tower 16	08:00:00	500	915	182.8%	3.65
17	Tower 17	16:00:00	1,012	1,570	155.1%	3.63
18	Tower 18	08:00:00	399	877	219.6%	3.56
19	Tower 19	08:00:00	2,394	4,571	190.9%	3.55
20	Tower 20	08:00:00	1,444	2,750	190.5%	3.54
21	Tower 21	08:00:00	2,128	3,471	163.1%	3.54
22	Tower 22	08:00:00	1,661	2,936	176.7%	3.53
23	Tower 23	16:00:00	2,826	4,094	144.9%	3.47
24	Tower 24	08:00:00	822	1,344	163.5%	3.45
25	Tower 25	08:00:00	459	1,019	221.9%	3.41

Event Report Details

The event report shows the details of the mass event.

The map and table show the clusters where most subscribers from the event were also seen that day.

The line chart shows the normal pattern of subscribers in this cluster - some clusters, such as those in market towns, may show a consistent weekday pattern.

The Mass Event Map provides a view of all the gatherings that occurred over a single day in Malawi. To facilitate users' understanding, the dashboard only displays the top 25 events on a given day, as arranged by z-score. Additional events can be displayed if necessary.

Three key metrics are highlighted:

- **The severity of the event** (the z-score)
- **The magnitude** (the number of people)
- **The relative magnitude** (the number of people as a percentage of the expected number of people)

In Figure 18, the map on the left shows locations, with the size of the dot indicating magnitude and the color indicating severity. The table shows the same information. Crucially, the magnitude is expressed as a relative measure to protect the proprietary data of the MNO, where knowing the exact number of users at each tower is valuable market information.

It is expected that users will scan the map for events that look plausible, then potentially investigate whether they occurred. Users can select any day to view historical data. Clicking on the map or a row on the table will bring users to the event detail section, as seen below.

Mass Event Detail View

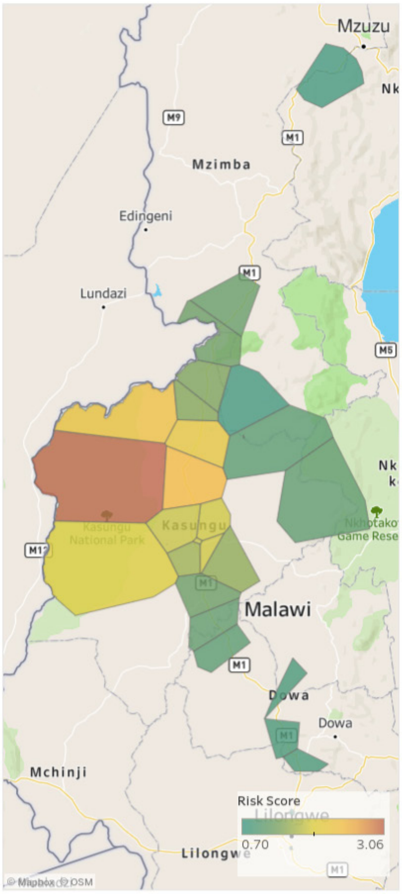
FIGURE 19

Detailed Event Report, August 2020

Note: Data shown in this image has been scrambled to protect confidential data shared by the MNO

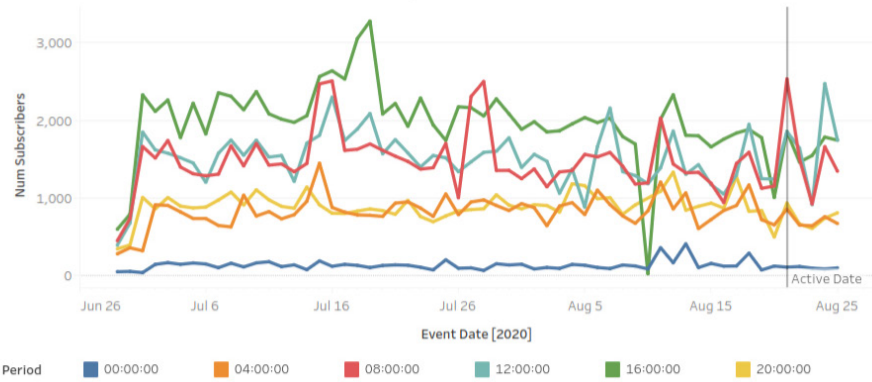
Event Report for Chulu Trading on August 21, 2020 at 08:00:00

Cluster Risk Map



1,229	300.0%	4.12
Total Subscribers	of Expected	Variance Score

Subscribers in Cluster over Last 60 Days



Highest Risk Clusters

Cluster ID	Cluster Name	TA	District	# Subs	Risk
68	Tower 1	Kasungu Boma	Kasungu	59	1.77
72	Tower 2	Kasungu Boma	Kasungu	54	1.73
110	Tower 3	TA Kaluluma	Kasungu	24	1.38
184	Tower 4	Mponela Urban	Dowa	7	0.85
225	Tower 5	TA Chakhaza	Dowa	10	1.00
263	Tower 6	Kasungu Boma	Kasungu	64	1.81
367	Tower 7	TA Mabulabo	Mzimba	6	0.78
388	Tower 8	Area 36	Lilongwe City	5	0.70
391	Tower 9	Kasungu Boma	Kasungu	34	1.53
396	Tower 10	TA Chilowamatambe	Kasungu	28	1.45
427	Tower 11	TA Njombwa	Kasungu	12	1.08
469	Tower 12	Area 51	Lilongwe City	6	0.78
512	Tower 13	TA Mponela	Dowa	7	0.85
525	Tower 14	TA Njombwa	Kasungu	16	1.20

The detailed event report displayed the risk profile of a single event. It enabled users to obtain additional information on the history of subscribers' movements in and out of that cluster over time. In Figure 19, the detail view situated the event in the history of that cluster (here from June 26 to August 25, 2020), and displayed key indicators (number of subscribers) and the risk spread profile (the other clusters people left and went to before and after going through that cluster).

Timelines showed the normal pattern of subscribers for that cluster, allowing the user to visualize just how anomalous an event was. The timelines helped users differentiate regular events (i.e., spikes that recur on a regular basis, as they would for a weekly market) from one-off anomalous gatherings. Knowledge of this pattern could enable a user to plan for next week's market or confirm whether markets were closing early per government mandate.

The risk map (left) showed the spread of risk for the clusters, as outlined above, defaulting to a view by TA. If the mass event was determined to be a danger, this showed users where a swift response may have been necessary.

SUSTAINABILITY AND SUPPORT TO GOVERNMENT

Iterative Development

Throughout the project phases, Cooper/Smith used an iterative and inclusive approach with the MoH to ensure product optimization and adoption. New products were quickly updated on a regular basis and presented to key stakeholders at the MoH and Public Health Institute to collect feedback and modify products to better meet country needs. This feedback loop has facilitated product adoption at all levels of the health system.

Stakeholders involved in this process included:

- Key PHIM personnel
- Digital Health Technical Working Group (70 participants)
- The MNO
- Surveillance sub-cluster of the COVID-19 response
- Emergency Operations Center meeting (all 23 COVID-19 clusters and sub-clusters)
- Health Cluster (attended by all clusters/sub-clusters, supporting partners, and COVID-19 Presidential Task Force members)
- MoH senior management team
- National Disaster Preparedness and Relief Committee
- Presidential Taskforce on COVID-19

Capacity Building

Within the Malawi MoH

The immediate next steps to ensure this model will continue to be supported include building systemic and organizational capacity to support the governance, management, and technical maintenance of the MNO data pipeline that will ensure its continuation.

In early 2021, Cooper/Smith launched a capacity-building plan[23] with the MoH that enabled its technical team to maintain and update the population mobility models based on MNO data analysis by the end of April.

---

[23] Capacity Building and Assessment Plan for MNO Data Analysis in Malawi: https://docs.google.com/spreadsheets/d/1jV0BCBkLshkcZUg0RT8z329BLjDrnNhaPtFAxezQFW0/edit#gid=641354308

The capacity-building plan's key components (systemic, organizational, and individual capacities) are outlined below:

System capacities:

- Introduce mobile data work
- Review data hosting plans
- Identify an MoH point of contact to begin a warm handover for relationships established during the DIAL Phase I and II

Organizational capacities:

- Implement an iterative process for key decision-makers and leaders
- Engage the MoH to review the existing DHD organization structure to ensure capacity exists for sustainability
- Establish a capacity-building plan
- Develop and implement responsible data use processes (DIAL 2021)

Individual capacities:

- Assess current skill level
- Develop individual capacity building plans
- Train staff on responsible data use
- Set up mentorship with MNO technical support

Institutionalization Across the Government of Malawi

Use cases in project Phases I and II have demonstrated the value of combining population mobility insights gained from mobile data analyses with other health-related datasets. Discussions with the Malawi Department of Disaster Management Affairs (DoDMA) have also shown the potential to leverage this kind of analytics to improve disaster management.

In Cooper/Smith and DIAL's view, long-term success of these initiatives will be achieved when—across all development sectors and geographies—data models have become:

- **Institutionalized** Data models are used, maintained, and updated by a governmental organization
- **Repeatable** The analytical process can be done over and over again and produce consistent results
- **Replicable** The analytical process can be used across sectors
- **Scalable** The data processing infrastructure can accommodate ever-growing data sets
- **Sustainable** The government can ensure the continuity of data access, data processing infrastructure, and data analysis

As such, Cooper/Smith and DIAL have advocated that other sectors within the GoM (e.g., education, agriculture) would benefit from the use of such analyses. The involvement of multiple stakeholders within the government would require building capacity outside of one specific ministry, ideally within a cross-sectional governmental agency.

Cooper/Smith and DIAL have engaged in discussions with several local stakeholders to define such a system. One first step would be to identify a governmental agency, independent from any ministry, that will become the custodian of the models and feed all branches of government with the novel data analytics. In Malawi, the NSO or the NPC could potentially play such a role.

MNO Data Access

Cooper/Smith received the data through agreements with the MNOs that were limited in time (duration) and scope (which data could be provided and how). As the MoH and other stakeholders in Malawi have expressed the desire to use MNO data analysis, new agreements will have to be established between MNOs and the GoM. Conscious of potentially limited funding available to the MoH, Cooper/Smith suggests setting up a co-creation value model with MNOs to enable the MoH to continue receiving CDR data at no cost.

A noncommercial, co-creation model would allow the MNO to share data with the MoH and Cooper/Smith in exchange for use of the data model for commercial purposes. The co-creation value model will give the MNOs access to Cooper/Smith's deep analytics expertise, while strengthening its customer segmentation and targeting capabilities, unlocking new commercial and marketing opportunities. This co-creation model would have two main parts:

- The MNO provides data to Cooper/Smith to feed the different models and support the work with the GoM
- Cooper/Smith makes the models available to the MNO and supports its effort to leverage the models for business intelligence

Cooper/Smith and the MoH of Malawi have engaged the MNOs to discuss the establishment of the co-creation model. As of April 2021, discussions are ongoing.

Data Hosting

Through project Phase I and Phase II, authorities in Malawi (IRB, MACRA, MoH) have authorized Cooper/Smith to store the data in the cloud, as detailed in the security infrastructure section (page 18). However, they have expressed the necessity of eventually having the data hosted in country. Cooper/Smith has worked with local stakeholders to identify a safe and reliable in-country data hosting solution and identified three potential options:

- **Co-sitting hosting:** The MoH owns the server, but it is located within the internet service provider (ISP) facility. The ISP provides access to internet and electricity and ensures server maintenance. One solution would be to have this co-sitting hosting solution set up with the MNO providing Cooper/Smith with the data.
- **Containerized data center:** This is a modular data center incorporated into a standard shipping container and includes all the necessary components used in a data center: cooling, power, racks, etc. The MoH has identified this means of hosting as a potential viable option, estimated the cost of bringing one to Malawi, and is currently looking for a donor to support this project. It has not received support yet.
- **National data center:** The GoM is building a national data center that will host all governmental data. However, the data center is still under construction.

For all three options, Cooper/Smith will have to continue to use the data architecture described in the security infrastructure section as it will still need AWS computing power to process the data for the modeling work.

As of April 2021, due to the lack of visibility of the MoH's preferred options (2 and 3), Cooper/Smith and the MoH have decided to keep the existing setup, with data hosted and processed on AWS. Both stakeholders will review the three options periodically and move to local hosting when one of the options becomes more concrete.

NEXT STEPS

Potential for Future Model Development

From its original development in March 2020, the epidemiological model has constantly refined and improved. In the first year of COVID-19, 75,000 peer-reviewed and preprint articles were written about COVID-19, with many developments and lessons learned along the way. Cooper/Smith continuously incorporated these findings into the model as they became available, including improvements in estimates of infection fatality rate, hospitalization rates, treatments, vaccines, genetic variants, and other novel discoveries.

Furthermore, as the use of CDR data becomes more common and methodological improvements are made to existing usage of the data, Cooper/Smith aims to incorporate those approaches as warranted. By further refining the model, there is potential to adapt the solution to support ongoing monitoring efforts within Malawi as the primary indicators for response may continue to change, while also presenting a critical opportunity to capture and produce insights for broader dissemination within the development community.

Replication

All the technical outputs produced through project Phase II (e.g., CDR data pipeline, CDR data processing, mass event detection and risk scoring, epidemiological model) have the potential to be reused in Malawi for other purposes outside of health. For example, Cooper/Smith has explored leveraging project outputs for disaster prevention. MNO data analysis can be leveraged to improve disaster response management, as demonstrated after the 2010 earthquake in Haiti.[24] Flooding is common during rainy season in Malawi, and some nonroutine events (e.g., hurricanes) often displace populations and put people at risk.

At present, Cooper/Smith has engaged the Malawi Department of Disaster Management Affairs (DoDMA) to explore a potential collaboration, applying the approach developed under this project in a disaster warning and response setting. While not formalized, the future development of these models has also been discussed with partners and donors to consider opportunities for applying the same methodologies and refinements to address future pandemic response efforts throughout sub-Saharan Africa and beyond.

---

[24] Bengtsson, L., Lu, X., Thorson, A., Garfield, R., von Schreeb, J. (2011). Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. *PLoS Medicine 8*(8): e1001083. https://doi.org/10.1371/journal.pmed.1001083

Long-Term Sustainability and Increasing Use

Though Phase II has ended, Cooper/Smith will continue engaging GoM stakeholders to explore additional use cases beyond the health sector in addition to institutionalization efforts to promote sustainability. Cooper/Smith and DIAL see the long-term success of these initiatives when, across all development sectors and geographies, data models have become *institutionalized, repeatable, replicable, scalable*, and *sustainable*.

Cooper/Smith and DIAL see the long-term success of these initiatives when, across all development sectors and geographies, data models have become *institutionalized, repeatable, replicable, scalable*, and *sustainable*.

Consequently, Cooper/Smith is also working with DIAL to conceptualize a potential future resource model within the GoM that would formally institutionalize these technical skills, models, and methodologies, while allowing other partner and donor organizations to invest directly in the development of response efforts rather than introducing external, short-term efforts. While a single entity within the GoM will likely be identified as the primary management and governing body, our approach will cut across all relevant GoM institutions. The operationalization of this model would involve support through tailored activities designed by Cooper/Smith, DIAL, and others, including, but not limited to, the following activities:

- **Refinement and institutionalization of key governance structures**, in alignment with GoM policies, priorities, and investments across donor and bilateral investments
- **Technical training of GoM staff** on the development, use, and maintenance of data models, predictive analytics, and other advanced analytics solutions for government use
- **Knowledge transfer to support the GoM** in evaluating the systems, infrastructure, hosting, and management needs of new data initiatives and models for national decision-making
- **Establishment of mechanisms to support the analysis, evaluation, refinement, and integration considerations of external/organizational data strategies**, ensuring alignment with GoM priorities prior to implementation
- **Financial resource requirement trainings** to ensure investments are adequately budgeted for within the GoM or clearly articulated to partners interested in leveraging GoM resources to support development projects and initiatives

With the additional, necessary support, Cooper/Smith and DIAL will work with the GoM to identify and train the most relevant governmental agency to serve as the central coordinating and management body for data models all relevant ministries. As such, multiple new use cases using mobile data analyses and capacity building would be at the heart of a potential project Phase III.

This initiative would require multiple years of investment and support, both on behalf of the GoM and external donors and partners. In the meantime, Cooper/Smith and DIAL will continue to collaborate and holistically leverage this concept, beginning with partnership discussions thought DoDMA as the first of many potential use cases.

LESSONS LEARNED AND RECOMMENDATIONS

Cooper/Smith, with the support of DIAL and other partners, presented a suite of analytics and tools utilizing MNO data that demonstrably improved decision-making for Malawi's pandemic response. The core analytics developed for this project—the epidemiological model and mass gathering event dashboard—can be used in tandem with traditional tools for COVID-19 surveillance and control, such as testing and tracing.

Use by stakeholders in Malawi demonstrated that the mass gathering event dashboard is a useful tool for identifying potential hotspots of COVID-19 transmission. Although COVID-19 testing in Malawi has increased tremendously since December 2020, it still lags behind the averages of other African countries by three- to 10-fold, and behind North America and Europe by 75-fold. Further, Malawi is not meeting either of the two WHO testing recommendations.[25] Cooper/Smith believes that leveraging the mass event dashboard could help direct scarce testing resources to areas where transmission is most likely to have occurred.[26]

Use by stakeholders in Malawi demonstrated that the mass gathering event dashboard is a useful tool for identifying potential hotspots of COVID-19 transmission.

Additionally, the epidemiological model, which was enhanced through the use of mobile phone data, is a critical tool in understanding the historical and future trajectory of the COVID-19 burden. This is especially critical during very intense periods of transmission, or "waves," when policymakers are interested in understanding the consequences of choices such as business closures, masking, physical distancing measures, school closures, vaccinations, and other approaches.

These tools are works in progress, and there are several areas that Cooper/Smith has identified for future improvement, most critically better understanding Malawi-specific COVID-19 data such as clinical characteristics, risk factors for severity, hospital and ICU use, and all-cause mortality data. Having information on these data points, especially at the district or subdistrict level, would provide valuable markers for calibrating and validating epidemiological models. These data may even be important for fully contextualizing and understanding current burden and health system strain, as well as urgency for policy response.

[25] 1,200 tests per 100,000 people over 28 days: https://www.cdc.gov/coronavirus/2019-ncov/travelers/how-level-is-determined.html
[26] Cooper/Smith Medium Blog, Part 1: How to use your data to fight COVID-19, a roadmap for countries in Sub-Saharan Africa: https://medium.com/cooper-smith/how-to-use-your-data-to-fight-covid-19-a-roadmap-for-countries-in-sub-saharan-africa-8e8b3967ce15

These novel analyses provide valuable insights for other, non-health-related use cases, as shown by the discussion with Malawi's DoDMA, which sees the value of such information to improve disaster management. Similar analyses could also be replicated to support other sectors, such as agriculture or education. As such, Cooper/Smith and DIAL aim to engage GoM officials to build capacity, possibly within a cross-agency, governmental organization that would serve all branches of government and provide these and related services through in-house technical support.

The global COVID-19 pandemic has taught public health officials the importance of rapidly analyzing trends and responding accordingly. During the most intense outbreaks of COVID-19, cases, hospitalizations, and deaths can double in a matter of days, quickly overrunning health system capacity. This leaves policymakers little time to understand the current situation, quantify trends, and weigh the costs and benefits of policy options. This project has demonstrated the potential of novel sources of data, like mobile phone records, to provide decision-makers in critical roles with readily accessible, rich analytics to support vital decisions that increase efficiency and save lives.

This project has demonstrated the potential of novel sources of data, like mobile phone records, to provide decision-makers in critical roles with readily accessible, rich analytics to support vital decisions that increase efficiency and save lives.

BIBLIOGRAPHY

American Hospital Association. (2020). *A compendium of models that predict the spread of COVID-19*. https://www.aha.org/guidesreports/2020-04-09-compendium-models-predict-spread-covid-19

Cooper/Smith. (2020). *Malawi-COVID-epidemiological-model: Repository to store files for COVID modeling*. GitHub. https://github.com/CooperSmith-org/Malawi-COVID-Epidemiological-Model

DIAL, Cooper/Smith, Infosys. (2019). *Using mobile phone data to make policy decisions: A study in how new data sources optimized health facility placement in Malawi.* https://digitalimpactalliance.org/research/using-mobile-phone-data-to-make-policy-decisions/

DIAL, Data Pop Alliance. (2021). Mobile data for development (MD4D) handbook, https://resources.dial.community/resources/md4d_handbook

GSMA. (2019). *The mobile gender gap report, 2019.* https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/02/GSMA-The-Mobile-Gender-Gap-Report-2019.pdf

McAloon, C., Collins, Á., Hunt, K., Barber, A., Byrne, A. W., Butler, F., Casey, M., Griffin, J., Lane, E., McEvoy, D., Wall, P., Green, M., O'Grady, L., & More, S. J. (2020). Incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. *BMJ Open*, *10*(8), e039652. https://doi.org/10.1136/bmjopen-2020-039652

Park, M., Cook, A. R., Lim, J. T., Sun, Y., & Dickens, B. L. (2020). A systematic review of COVID-19 epidemiology based on current evidence. *Journal of Clinical Medicine, 9*(4), 967. https://doi.org/10.3390/jcm9040967

Rees, E. M., Nightingale, E. S., Jafari, Y., Waterlow, N. R., Clifford, S., Carl, C. A., Group, C. W., Jombart, T., Procter, S. R., & Knight, G. M. (2020). COVID-19 length of hospital stay: A systematic review and data synthesis. In *BMC Medicine, 18*(1), 270). BioMed Central Ltd. https://doi.org/10.1186/s12916-020-01726-3

Reiner, R. C., Barber, R. M., Collins, J. K., Zheng, P., Adolph, C., Albright, J., Antony, C. M., Aravkin, A. Y., Bachmeier, S. D., Bang-Jensen, B., Bannick, M. S., Bloom, S., Carter, A., Castro, E., Causey, K., Chakrabarti, S., Charlson, F. J., Cogen, R. M., Combs, E., … Murray, C. J. L. (2021). Modeling COVID-19 scenarios for the United States. *Nature Medicine*, *27*(1), 94–105. https://doi.org/10.1038/s41591-020-1132-9

UNOCHA. (2020). *Peer review framework for predictive analytics in humanitarian response.* https://data.humdata.org/dataset/2048a947-5714-4220-905b-e662cbcd14c8/resource/76e488d9-b69d-41bd-927c-116d633bac7b/download/peer-review-framework-2020.pdf

Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P. G. T., Fu, H., Dighe, A., Griffin, J. T., Baguelin, M., Bhatia, S., Boonyasiri, A., Cori, A., Cucunubá, Z., FitzJohn, R., Gaythorpe, K., … Ferguson, N. M. (2020). Estimates of the severity of coronavirus disease 2019: A model-based analysis. *The Lancet Infectious Diseases, 20*(6), 669–677. https://doi.org/10.1016/S1473-3099(20)30243-7

Walker, P. G. T., Whittaker, C., Watson, O. J., Baguelin, M., Winskill, P., Hamlet, A., Djafaara, B. A., Cucunubá, Z., Mesa, D. O., Green, W., Thompson, H., Nayagam, S., Ainslie, K. E. C., Bhatia, S., Bhatt, S., Boonyasiri, A., Boyd, O., Brazeau, N. F., Cattarino, L., … Ghani, A. C. (2020). The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science, 369*(6502), 413–422. https://doi.org/10.1126/science.abc0035

APPENDICES

Appendix A - Co-creation Value Concept Note

DIAL and Cooper/Smith (the "Project Team") understand that achieving sustainability in this partnership depends on the Project Team's efforts being aligned with the MNO's commercial realities. The proposed business model calls for a self-sustained instantiation with all proceedings from insights (net of technical costs) available to the MNO. The Project Team recommends a co-creation value, nonfinancial model in which:

- The MNO provides data to the Project Team to feed the different models and support the work with Malawi MoH.
- Cooper/Smith makes the models available to the MNO and supports its effort to leverage the models for business intelligence.

Note that due to the exploratory nature of this project, the Project Team will work with the MNO to determine the correct technical, operational, and commercial models to ensure success.

Through this project, the Project Team can bring value to the MNO in two main ways:

Make deep analytics expertise available:
Cooper/Smith is a highly specialized data management firm with unique capabilities to leverage large data sets and transform data into insights. Cooper/Smith will replicate this approach with the MNO. It will share the models and other deliverables built through this project with Airtel and work with the company to support them and adapt the models to their needs.

Create a more granular customer segmentation to improve customer targeting:

a. **Demographics**
   Using the developed model and adding other data, Cooper/Smith will be able to create a customer segmentation by:
   - Income
   - Behavior
   - Attitudes/beliefs/perceptions

b. **Location and movement**
   Combining Airtel data with population data (e.g., from the census, WorldPop, or satellite data), Cooper/Smith will triangulate users and be able to determine patterns in population movement that will help the MNO understand customer movements and how to engage optimally

c. **Social networks**
   Using the model over time and with the inclusion of real-time data, Cooper/Smith will first identify network types (family, school, professional, etc.) between users and determine social behavior patterns (e.g., how people behave during a holiday)

Appendix B - Additional Information on SEIR Model

Model Development

The model described here is a dynamic, compartmental, deterministic susceptible-exposed-infected-recovered (SEIR) model, which simulates contacts between individuals daily, estimating the number of transmissions. It also estimates the number of infected persons who would need either hospitalization or more significant intensive care, or who are dying. The model is age-structured (<20 years, 20-49 years, and 50+ years) with differential age-mixing rates. The model is parameterized with estimates of transmission (the basic reproductive number – R_0), as well as rates of recovery, hospitalization, intensive care need, and death (*Figure 20*). Note that direct, empirical data on clinical characteristics of COVID-19 patients from African nations, especially low-income nations such as Malawi, have been extremely limited. Cooper/Smith prioritized high-quality estimates from Malawian and African settings when considering parameter value choice. When local data was not available, Cooper/Smith ensured that estimates were standardized and adjusted relative to the age structure of Malawi, which is significantly younger than contexts in North America and Europe, where much clinical research on COVID-19 is sourced.

Critically, calibration of the mathematical model in the Malawian context has been challenging. Since the start of the epidemic, Malawi has performed a cumulative 76 tests per 100,000, which is three times lower than the African average of 220 per 100,000. In terms of testing trends, from April to December 2020, the average number of daily tests was about 275, or 1.4 per 100,000 residents. Daily tests have increased dramatically throughout December 2020 and January 2021, to a high of 3,000 tests per day, or 14 per 100,000. Despite this acceleration, the number of tests still lags behind international standards by a factor of 20 times, with more than 300 daily tests per 100,000 residents in Europe and North America over the same period. Even with the much greater testing rates in high-income settings, underdiagnosis rates have been estimated between three- to 10-fold. In the case of Malawi, this suggests an even more extreme rate of under-ascertainment of infections. Calibrating the model to official caseloads would result in significantly underestimating the burden and spread of COVID-19. Throughout the epidemic in Malawi, Cooper/Smith estimated that the rate of underdiagnosis has ranged from a low of 50-times in December 2020 and January 2021 to as high as 500-times from May to September 2020.

FIGURE 20

Epidemiological Model Key Parameters, Values, and Sources

Parameter	Value	Source
R0	1.9	(Park et al., 2020)
Infectious time (days)	7	(McAloon et al., 2020)
Hospitalized time (days)	4	(Rees et al., 202
ICU time (days)	8	
Hospitalized rate of infected		
Pediatrics (<20)	0.009%	(Verity et al., 2020)
Adults (20-49)	1.2%	
Elderly (50+)	5.5%	
ICU risk among hospitalized		
Pediatrics (<20)	5.0%	(Verity et al., 2020)
Adults (20-49)	14%	
Elderly (50+)	28%	
Fatality rate of ICU		
Pediatrics (<20)	9%	Verity et al., 2020)
Adults (20-49)	20%	
Elderly (50+)	59%	

Mathematical models in other settings have calibrated disease dynamics relative to mortality, or in some cases “excess mortality,” an estimate of the true mortality of COVID-19 that uses historical rates of total mortality as a reference point to estimate the unmeasured COVID-19 deaths due to limited testing. In contexts like the United States and Europe, where mortality trends are well defined and data are rapidly reported, this process has been easily incorporated into models. This effect can be quite large, estimated as high as 100% above measured COVID-19 deaths in heavily hit countries such as Italy, the United States, the United Kingdom, and Spain.

However, this technical approach was not a reasonable option in Malawi, where fewer than one in six deaths of any cause are reported and the reporting may be delayed by months. There are many reasons to believe COVID-19 mortality in Malawi is less than other nations, primarily due to its significantly younger median age of 19 years. However, there are many unknowns, such as the effects of common conditions such as HIV, malaria, tuberculosis, and malnutrition on COVID-19 severity. Indeed, studies have indicated that the risk of death could be doubled among those with HIV, TB, or malnutrition.[27] These studies have been few, small in sample size, and imperfect in design and, therefore, make no further adjustments to severity of disease estimates beyond adjusting for age.

Furthermore, the geospatial distribution of the few tests conducted (and their positivity rate) has not been well described. Still, there is ample evidence that there is considerable heterogeneity in geographic epidemic timing and characteristics related to population density, demographics, health care access, and background risk of severity. Therefore, Cooper/Smith sought to adjust our model to account for differences in epidemic trajectories at the subnational level, specifically down to the TA.

Model Refinement

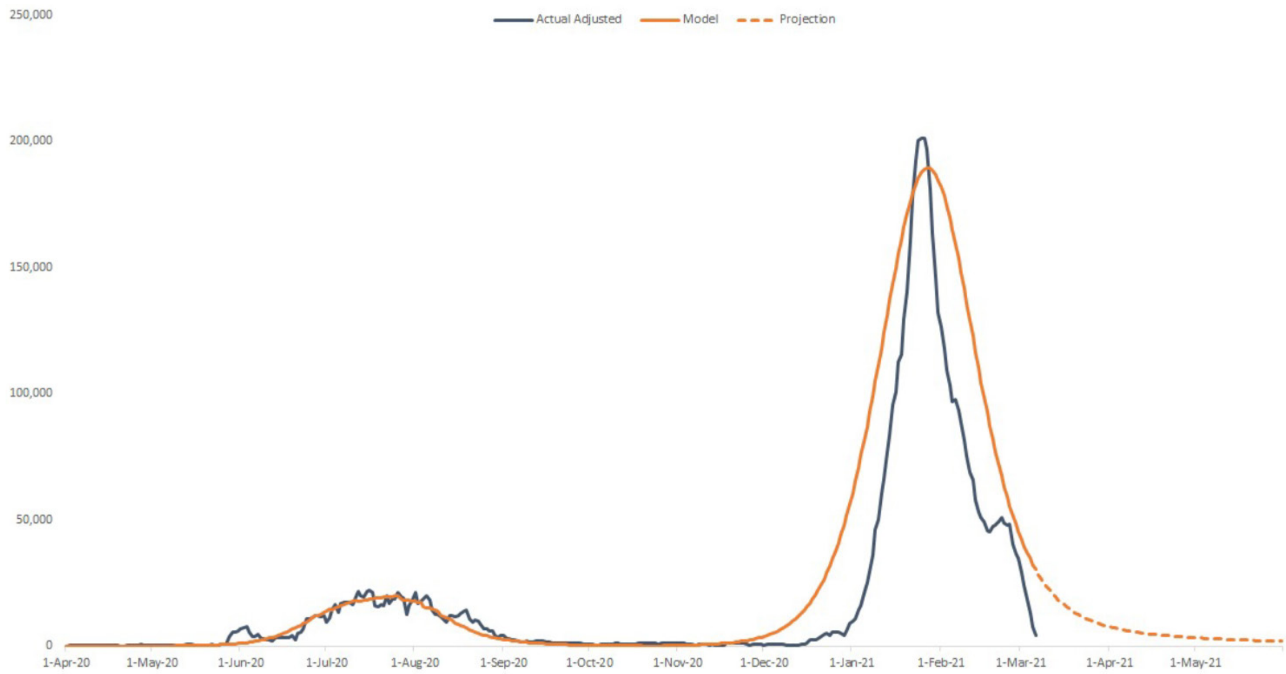
The mathematical modeling of COVID-19 in Malawi was an ongoing process in which Cooper/Smith constantly reviewed and considered the latest epidemiological and clinical data. Notably, in late 2020 there were two major developments that are related to one another. First, many African nations experienced an exponential growth in COVID-19 cases and deaths following a lengthy quiet period of the epidemic. This may partially be explained by relaxation of policies, enforcement, or adherence among citizens; seasonal changes in transmissibility; and/or increases in international travel and higher likelihood for mass importation of infections. The latter could have been facilitated by repatriation of Malawian citizens from neighboring countries throughout southern Africa. Secondly, the emergence of several new genetic variants of COVID-19 became known, namely in the United Kingdom, South Africa, and Brazil, although it is uncertain whether the underlying mutations occurred in those nations. Nonetheless, each are believed to be significantly more transmissible (as high as 50% greater) and may evade previous immunity. As of February 2021, these variants have been known to become dominant in many settings and have been linked to tremendous COVID-19 resurgences in several settings.

The model estimated a cumulative 2.5 million infections through the end of January 2021 (*Figure 21*). Total diagnoses through this period were estimated to be 22,000. However, as previously described,this was certainly an underestimate of true infections. Therefore, Cooper/Smith adjusted that estimate relative to a rate of underdiagnosis that is linked to the per capita testing rate. This adjusted estimate of total infections was 2.1 million through the same period.

[27] Clinical Infectious Diseases. 2020 August 29; ciaa1198

FIGURE 21

Comparison of Modeled Infections and Adjusted Diagnosis in Malawi by Day



The model estimated a need of 1,600 hospitalizations and of 500 intensive care beds (*Figure 22*). Cooper/Smith further estimated a cumulative total of 1,500 COVID-19 deaths in Malawi through the end of January 2021 (*Figure 23*). This is three-fold greater than the official figure of 550. Among these deaths, Cooper/Smith estimated that 84% occur among those aged 50 years and up, 16% among those aged 20-49 years, and only 0.02% among those aged less than 20 (*Figure 24*).

FIGURE 22

Modeled Current Hospitalization and Intensive Care Need by Day

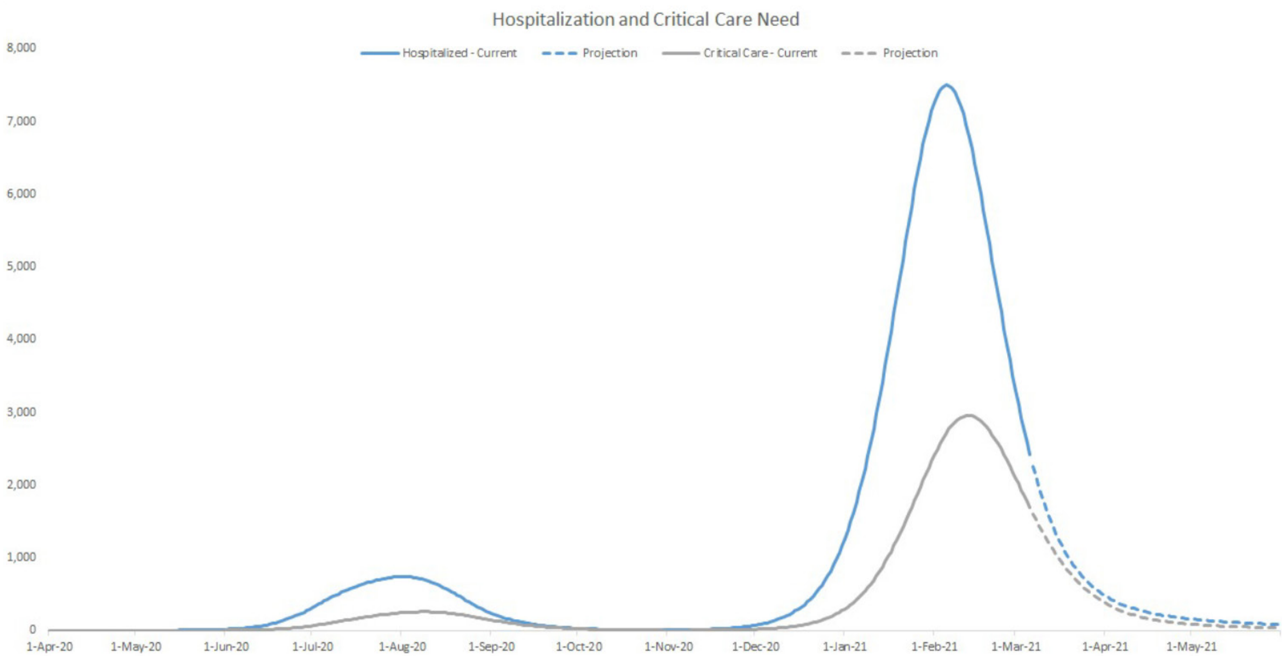


FIGURE 23

Modeled Cumulative COVID-19 Deaths by Day

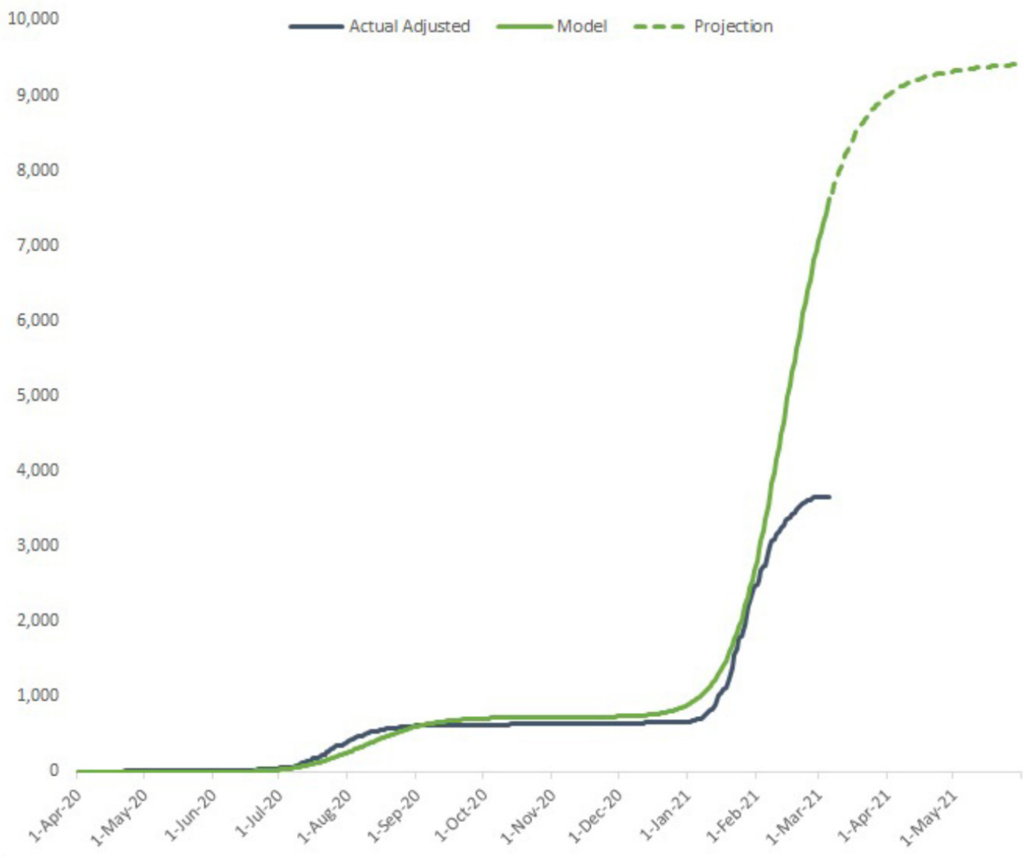
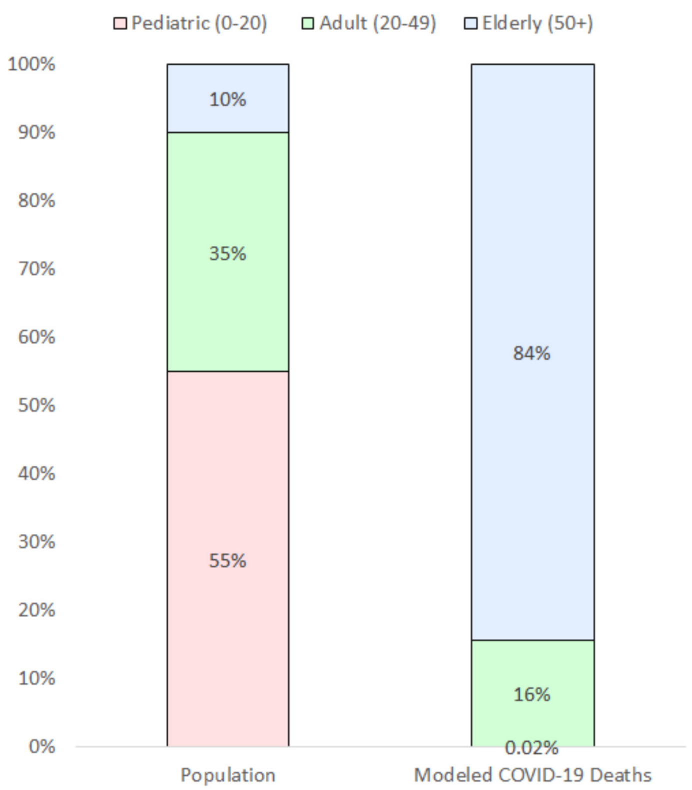


FIGURE 24

Distribution of Population and Modeled COVID-19 Deaths by Age



Appendix C - Description of COVID-19 Risk and Vulnerability Models

1) Risk Model

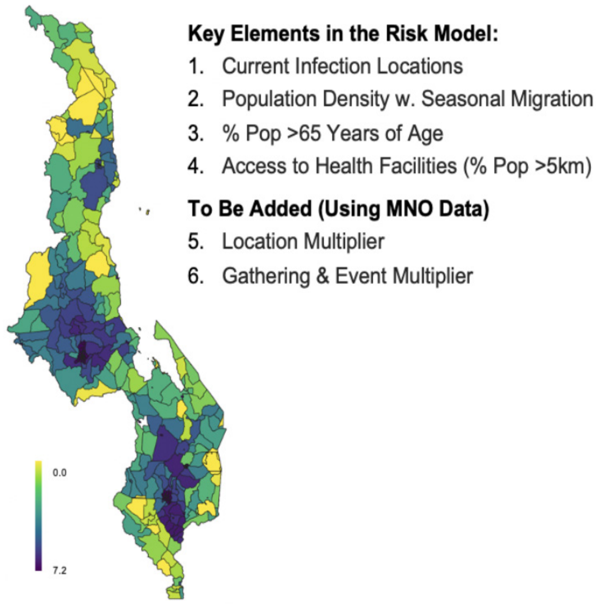
Cooper/Smith developed a risk model to help key stakeholders in the COVID-19 response (MoH national level; district health officers) identify high-need areas. The model consisted of risk factors inherent with COVID-19 (e.g., population density, percentage of population >60 years).

These factors were then brought together to develop a consolidated risk score at the lowest available administrative level, as can be seen in Figure 25. The darker the area, the more at risk the area is.

Cooper/Smith worked with the West African Health Organization (WAHO) to create a West Africa-wide risk model, enabling comparison between the 14 countries of the Economic Community of West African States (ECOWAS). Cooper/Smith also made the risk model accessible free of charge online[28] and supported users to replicate model use for their own country via a webinar and video tutorials.[29]

FIGURE 25

COVID-19 Risk Model for Malawi, May 2020

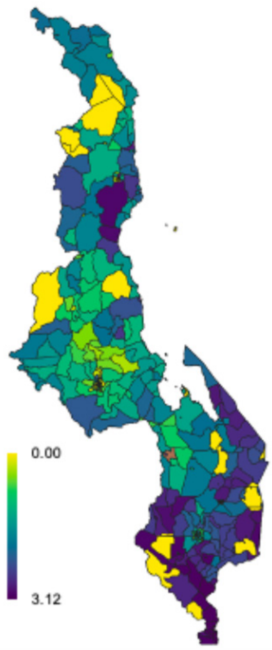


2) Vulnerability Model

Cooper/Smith also developed an economic vulnerability model to help the MoH understand which areas are most at risk of the social and economic negative impacts of COVID-19. The final ranking includes data on poverty levels, food security, hygiene (access to water and soap), and wasting and stunting prevalence, as can be seen in Figure 26. The darker the area, the more vulnerable the population is.

FIGURE 26

COVID-19 Economic Vulnerability Model for Malawi – July 2020



---

[28] Risk model available on GitHub: https://github.com/CooperSmith-org/ssa-risk-models
[29] Additional information available at https://coopersmith.org/covid-19

Digital Impact Alliance

digitalimpactalliance.org

COOPER/SMITH

www.coopersmith.org

UNITY AND FREEDOM

www.health.gov.mw