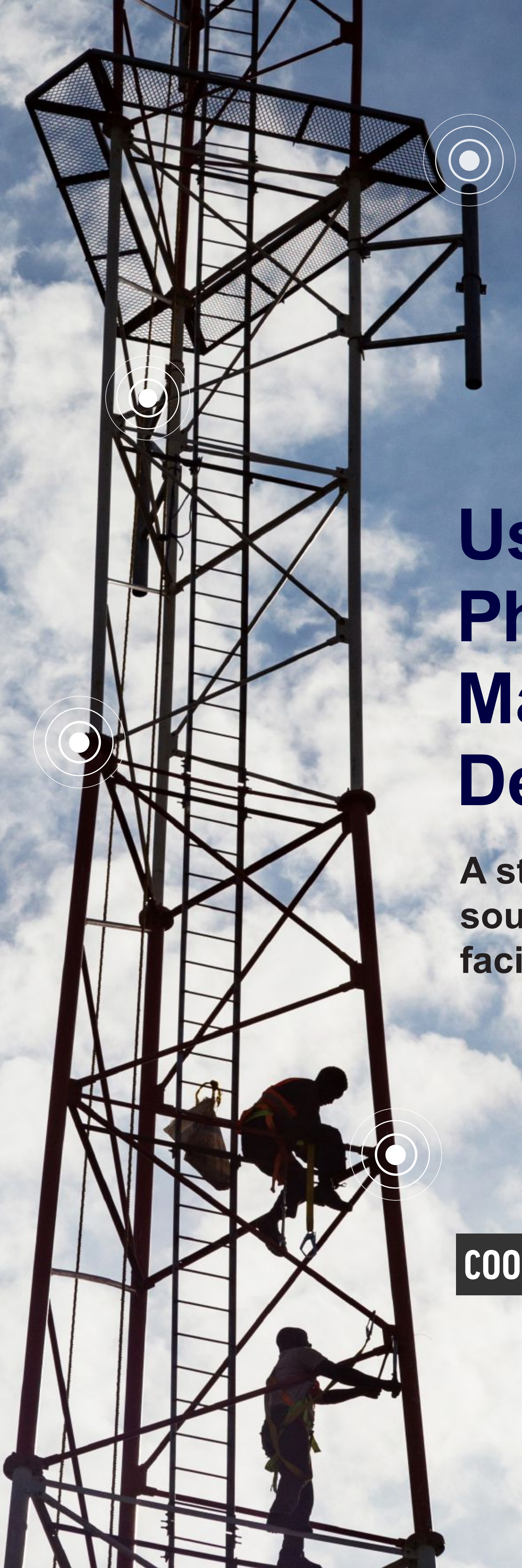


Using Mobile Phone Data to Make Policy Decisions

A study in how new data sources optimized health facility placement in Malawi

COOPER/SMITH

dial Digital Impact Alliance

Infosys®

ACKNOWLEDGMENTS

This technical report was prepared as part of a partnership between the Digital Impact Alliance (DIAL), Cooper/Smith, and Infosys Limited. Primary contributing authors include Erwin Knippenberg, Rachel Sibande, Emily Chirwa, Tyler Smith, Senthil Kumar Subramanian, Arbind Kumar Sinha, and Syed Raza. Special thanks are due to other members of the DIAL team, including Cristen Bauer, Sabrina Kang, and Nicholas Gates, as well as partner organizations like UNICEF and WorldPop, for data and technical support.

ABOUT DIAL

DIAL aims to realize a more inclusive digital society in emerging markets, in which all women, men and children benefit from life-enhancing, mobile-based digital services. A partnership among USAID, the Bill & Melinda Gates Foundation, the Swedish government, and the United Nations Foundation, DIAL helps accelerate the collective efforts of government, industry and NGOs to realize this vision.

DIAL is staffed by a global team and is guided by a board of leading emerging market entrepreneurs, technologists and development experts. With this leadership, DIAL is uniquely positioned to serve as a neutral broker, bringing together government, industry and other development stakeholders to promote new solutions to old problems. For more information about the Digital Impact Alliance or this technical report, please visit our website: www.digitalimpactalliance.org.

ABOUT COOPER/SMITH

Cooper/Smith is a technical assistance organization that uses hard data to increase the effectiveness and efficiency of development programs worldwide. The organization brings deep knowledge and expertise in strategic planning; health financing and evidence-based resource allocation; digital health systems implementation; operations, behavioral, and clinical research; as well as application of advanced analytic methods, including econometrics, modeling, and machine learning.

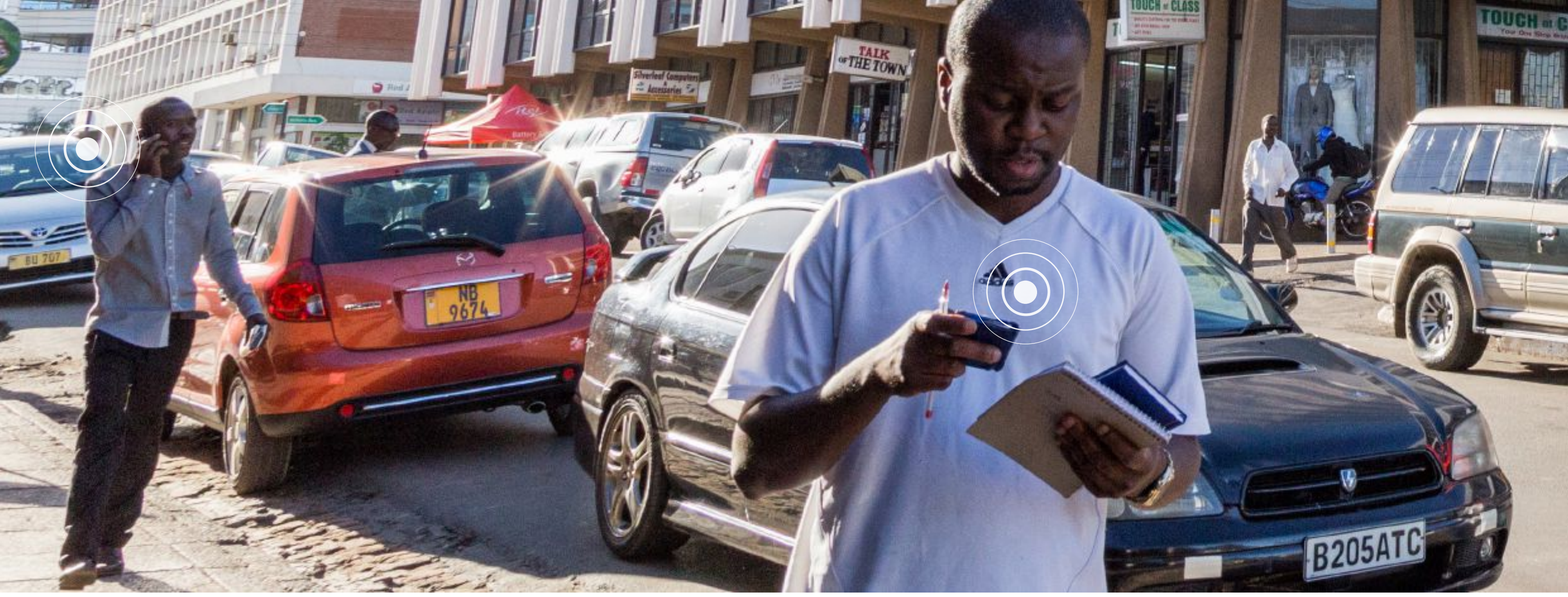
We support large-scale data systems and data use projects in Malawi, funded by the Bill & Melinda Gates Foundation and the Digital Impact Alliance. We support other programs globally, and locally within Burkina Faso, Cameroon, Kenya, Liberia, Madagascar, Nigeria, Senegal, Thailand, and Zambia. Former and current partners include Georgetown University; the London School of Hygiene and Tropical Medicine; the Boston Consulting Group; Catholic Relief Services; Canada's Minister of International Development; the Center for Strategic and International Studies; the International Fund for Agricultural Development; the Global Fund to Fight AIDS, Tuberculosis, and Malaria; the World Bank, and the Global Financing Facility in Support of Every Woman, Every Child. We participate in numerous technical and advisory groups and are committed to the democratization of data access and use worldwide.

For more information, please visit our website at www.coopersmith.org.

ABOUT INFOSYS

Infosys is a global leader in next-generation digital services and consulting. We enable clients in 45 countries to navigate their digital transformation. With over three decades of experience in managing the systems and workings of global enterprises, we expertly steer our clients through their digital journey. We do it by enabling the enterprise with an AI-powered core that helps prioritize the execution of change. We also empower the business with agile digital at scale to deliver unprecedented levels of performance and customer delight. Our always-on learning agenda drives their continuous improvement through building and transferring digital skills, expertise, and ideas from our innovation ecosystem. For more information, please visit our website at www.infosys.com.

TABLE OF CONTENTS



EXECUTIVE SUMMARY

In 2017, the Digital Impact Alliance (DIAL), Cooper/Smith, and Infosys (the Project Team) embarked on a collaborative effort with the Malawi Ministry of Health (MoH) and Malawi Communications Regulatory Authority to demonstrate how Mobile Network Operator (MNO) data analytics can be used to achieve development goals. The Project Team's working hypothesis was that analytics from MNO data could be used to complement other traditional datasets, such as census data, in order to inform policy and decision-making for governments and their agencies. This technical report discusses the findings from the Project Team's test case in Malawi where MNO data analytics were used to understand population density, migration, and urbanization patterns to help inform health facility planning.

The problem and use cases

The MoH launched a Capital Investment Plan, which included a proposal to build 900 new health posts across the country between 2020 and 2023. Currently, roughly half of the population in Malawi lives more than 5 km from a health facility, severely hampering their access to essential health care services. One of the primary objectives of the Capital Investment Plan was to strategically identify locations for these new health posts, such that it would ensure 95 percent of the population would be living within 5 to 6 km of a health facility by 2023. To that end, the MoH solicited input from the Project Team to help develop an analytics model to produce dynamic maps with optimal locations for the new health posts across Malawi.

Data collection, analysis, and validation model

The model used a combination of previously untapped analytics from MNO data and other analogue datasets in order to understand population growth, density, and migration in Malawi and project optimal placement of new health facilities in line with Malawi's national priorities. To develop this model, the Project Team collected and analyzed several datasets, including:

- Anonymized MNO data provided by an MNO partner, including call data records and geo-tagged location of cell towers;
- High-resolution population density data for 2015 that was compiled by WorldPop 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 MoH from a detailed UNICEF survey of every operating health facility in Malawi; and
- Monthly facilities-level disease burden data also provided by the MoH.

Using the data analytics to optimize clinic allocation

Using these key datasets, the Project Team was able to provide the MoH with a national view of health facility coverage against population density and movement patterns. These optimized allocation models were compared to the MoH's Capital Investment Plan criteria for the new health posts, while additional considerations were taken into account including current areas with an insufficient number of health posts, migration patterns, and vulnerable regions that periodically experience cut-off access to health facilities due to flooding.

Once the Project Team had predictions about population densities and population movement over time, the team validated these predictions by comparing them to the 2018 census data. The Malawi census data provided a detailed headcount of the population based on on-the-ground surveys at the national, district and traditional authority levels in 2018. The Project Team found there was less than a 5 per cent discrepancy between the predicted and actual population levels for each district, revealing that—in Malawi—satellite data and MNO data analytics are both good proxies for actual population levels. Final recommendations were then given to the MoH to inform the locations of new health posts and help them achieve their goal by 2023.

Key takeaways and recommendations

The Project Team's process of acquiring the data, analyzing it and integrating it into country systems generated a number of key findings, limitations, and recommendations.

Key Findings

1. An estimated 7.74 million Malawians, or 44.7%, live more than 5 km walking distance from a clinic or health post.
2. Observations from the deidentified MNO data analytics suggests there is significant population movement on weekends and during the rainy season.
3. The number of people serviced per new health post, a measure of allocation efficiency, varies significantly across districts.

Limitations

1. The model is gender-blind by construct, since the data was stripped of identifying characteristics.
2. Certain populations in other vulnerable groups such as the poor, the elderly, and children might also have different rates of mobile use.
3. The model is based on data from only one of two principal telecom providers in the country.
4. Nighttime location in deidentified MNO data is assumed to indicate where a person lives.
5. There is an unknown composition of cell phone users relative to those who tend to migrate; therefore, short- and medium-term population movements may be either overstated or understated.
6. By construct, there is no MNO data for the approximately 5% of zones outside of mobile coverage in Malawi.

Recommendations

1. Define a specific, demand-driven use case.
2. Bring together a broad-based analytical team of researchers.
3. Emphasize country-level buy-in from the very beginning of the project.
4. Engage with private-sector partners throughout the process.
5. Work continuously with technical counterparts.
6. Consider the country-specific data limitations.
7. Validate the model using robust census data.
8. Verify the performance of the model.

Sustainability and future work

This Malawi test case demonstrated a strong use case for how data for development projects can be deployed. This technical report provides evidence for the Project Team's working hypothesis that analytics from MNO data can be used to complement traditional datasets, such as census data, in order to inform policy and decision-making for governments and their agencies. Insights on population density and movement patterns might also be used in future projects to inform decision-makers about other key locations such as water points, schools, and agricultural cooperatives, in order to improve people's access to those critical services.

To ensure sustainability and replicability, the Project Team engaged with Malawi government partners to integrate the analysis into country systems and provide a ranked list of priority locations for new health posts. DIAL will continue to work with government partners in Malawi and other governments across the globe to illustrate how such data for development models can be made replicable across development sectors and geographies.

BACKGROUND

New data sources such as mobile phones and geographic information systems (GIS) are being widely used in the developed world for commercial and public service purposes. While developing countries have also begun to experiment with these new data sources, they have not yet been incorporated as routine practice in development situations where they could provide a public benefit. The Digital Impact Alliance (DIAL) and its partners are working to increase the use of these new types of data in the developing world for humanitarian purposes by identifying use cases across health, agriculture, education, and other sectors.

In 2017, the Digital Impact Alliance (DIAL), Cooper/Smith, and Infosys (the Project Team) embarked on a collaborative effort with the Malawi Ministry of Health (MoH) and Malawi Communications Regulatory Authority to demonstrate how Mobile Network Operator (MNO) data can be used to help deliver development gains. The Project Team's working hypothesis was that MNO data analytics could be used to complement other traditional datasets, such as census data, in order to inform policy and decision-making for governments and their agencies. This technical report discusses the findings from the Project Team's test case in Malawi where MNO data was used to examine population density, migration, and urbanization patterns to help inform health facility planning.

This technical report discusses the findings from the Project Team's test case in Malawi where MNO data was used to examine population density, migration, and urbanization patterns to help inform health facility planning.

The MoH solicited input from the Project Team for its Capital Investment Plan, and this initiative was launched following the release of that plan in Malawi, which included key targets for improving citizens' access to health services. In Malawi, an estimated 7.73 million people, or 44.7% of the population, live more than 5 km from a health facility, severely hampering access to essential healthcare services. Malawi's population is projected to grow from 17.6 million in 2018 to 21.6 million in 2023. Without action, the number of Malawians without access to a health facility would increase to 9.7 million by 2023.

One of the objectives included in the Capital Investment Plan was to identify the optimal location of these new health posts, to effectively ensure that 95% of the population would live within 5 to 6 km of a health facility by 2023. Additionally, the Plan prioritized approaches to increase resilience during the rainy season by allocating more health posts to areas where flooding might prevent access or regions significantly impacted by migration to during unstable weather periods. In order to more accurately capture population density and movement patterns throughout the country, the MoH solicited input from the Project Team to specifically inform their approach to building 900 new health posts between 2020 and 2023.

In order to more accurately capture population density and movement patterns throughout the country, the MoH solicited input from the Project Team to specifically inform their approach to building 900 new health posts between 2020 and 2023.

By definition, a health post is comprised of a facility staffed by a clinician or community health worker, stocked with essential medicines in order to provide frontline care. While the Capital Investment Plan also includes investments in secondary and tertiary care, the government emphasized its priority to expand access to frontline, primary healthcare facilities. The government and Project Team agreed that the targeted placement of these health posts was an important step towards providing widespread access to primary healthcare. These new health posts included both upgrades to existing facilities and the construction of new buildings to expand access to rural and remote areas.

In order to support the MoH and optimize the placement of these new health posts, the Project Team used anonymized, aggregated MNO data along with other traditional datasets, as a proxy to understand population densities, migration, and urbanization patterns in Malawi. The deidentified MNO data complements national census data, allowing for the dynamic tracking of population movement in a timely and cost-effective manner. The analysis from the Project Team was used to develop a model that was able to effectively forecast migration patterns and make projections about where to optimally place health facilities in line with Malawi's national priorities. For the purposes of this project, access to any health facility was considered sufficient, with health posts being considered as the most cost-effective means to expand access to primary health facilities.

METHODOLOGY

The Project Team developed innovative new methods and processes tailored specifically to the needs and goals outlined by the Malawi MoH. There were no pre-existing roadmaps or plans that fit the exact specifications or goals for this project. The methodology and analytical processes were tailor-made and adapted to accommodate the specific actors involved and customized based on Malawi's country context. In order to develop a custom methodology for this project, the Project Team pulled from a body of existing research from other institutions on population density, population movement, and using MNO data in development outcomes.

In order to develop a custom methodology for this project, the Project Team pulled from a body of existing research from other institutions on population density, population movement, and using MNO data in development outcomes.

Population density

With the increased availability of high-resolution satellite data combined with powerful algorithms, there have been sustained improvements in forecasting population density (Stevens et al., 2015). The WorldPop project, an open source collaboration between researchers, trains its algorithms on historical census data and uses those algorithms to project annual population density at a 100-meter resolution (WorldPop, 2018). This resource was used by the Project Team to evaluate trends and patterns in population density throughout the country of Malawi.

Population movement

Complementary to the population density mapping approach, researchers have demonstrated the potential to harness mobile phone data to map population movements dynamically. Deville et al. (2014) showed that the density of unique users in a cell tower's catchment area corresponded closely with population density. Therefore, researchers can extrapolate from this mobile phone data to predict shifts in population densities between day and night, weekdays and weekends, and across seasons. Using case studies conducted in Portugal and France, the researchers showcased significant shifts in population density, enough to impact service provision (Deville et al., 2014). Erbach-Schoenberg et al. (2016) showed how similar methods can be used in the context of developing countries to account for seasonal fluctuations in calculating disease incidence. Working with mobile phone data in Namibia, the researchers highlight how seasonal mobility affects estimates of malaria incidence, leading to differences of up to 30% when compared to estimates created using only static population maps.

MNO data and development outcomes

Finally, research focusing on service provision has sought to link mobile phone data to development outcomes. Blumenstock, Cadamuro, and On (2015) used mobile network data and machine learning algorithms to predict poverty outcomes in Rwanda, identifying hotspots with a high degree of precision. In the health sector, Wesolowski et al. (2015) demonstrated the relationship between travel distance, as calculated using a household's radius of gyration,[1] and the percentage of households not receiving antenatal care. Travel distance remains a salient issue when it comes to the provision of basic health services.

This report illustrates how the Project Team collaborated to develop an innovative policy-relevant use case that harnesses MNO data to identify current and future gaps in the availability of health services in Malawi.

This technical report represents a new approach to data for development that was inspired by the insights and methods from the above resources, literature, and studies. This report illustrates how the Project Team collaborated to develop an innovative policy-relevant use case that harnesses MNO data to identify current and future gaps in the availability of health services in Malawi. Working closely with the MoH, the Project Team drew up recommendations for where to optimally place new health posts in order to remedy those health service gaps.

[1] The radius of gyration is the distance from a central point defined as the center of inertia. In this context, it is the assumed distance a household can travel on average given its pattern of locations identified using MNO data.

DATA COLLECTION

Identifying Data Sources

The Project Team drew from existing literature to develop an analytical plan that sought to combine three streams of data:

1. Anonymized MNO data provided by an MNO partner, including:
   a. CDR data (2016-2017)
   b. Geo-tagged location of each cell phone tower
2. High-resolution population density data for 2015 compiled by WorldPop[2]
3. Location and catchment area of existing health facilities, provided by MoH in collaboration with UNICEF

The analysis proposed to map the deidentified MNO data to administrative units, calculate the density of unique callers, then calibrate this against population density using WorldPop data. Fluctuations in the density of callers can then be extrapolated to infer population migration trends.

Establishing Data Pipeline

In parallel, the Project Team worked to establish a viable pipeline to import MNO data to a secure in-country server. To ensure proper anonymization of the MNO data, Infosys provided training to MNO technical staff.[3] A total of 26 months of data representing every call and SMS sent and received in Malawi between January 2016 and May 2018 were transferred to a secure in-country server.[4] Access to the anonymized raw data was restricted to the server system administrator, with a small group of additional technical users given access only to the processed data.

Once stored on the secure server, the data was cleaned for analysis. In the process of cleaning, the Project Team flagged several issues concerning data quality, including missing observations. Addressing these concerns around data quality and preparing the data for analysis proved to take more effort and time than initially anticipated. The Project Team had to iterate several times with MNO counterparts to ensure the proper transfer and formatting of complete datasets, leading to delays.

---

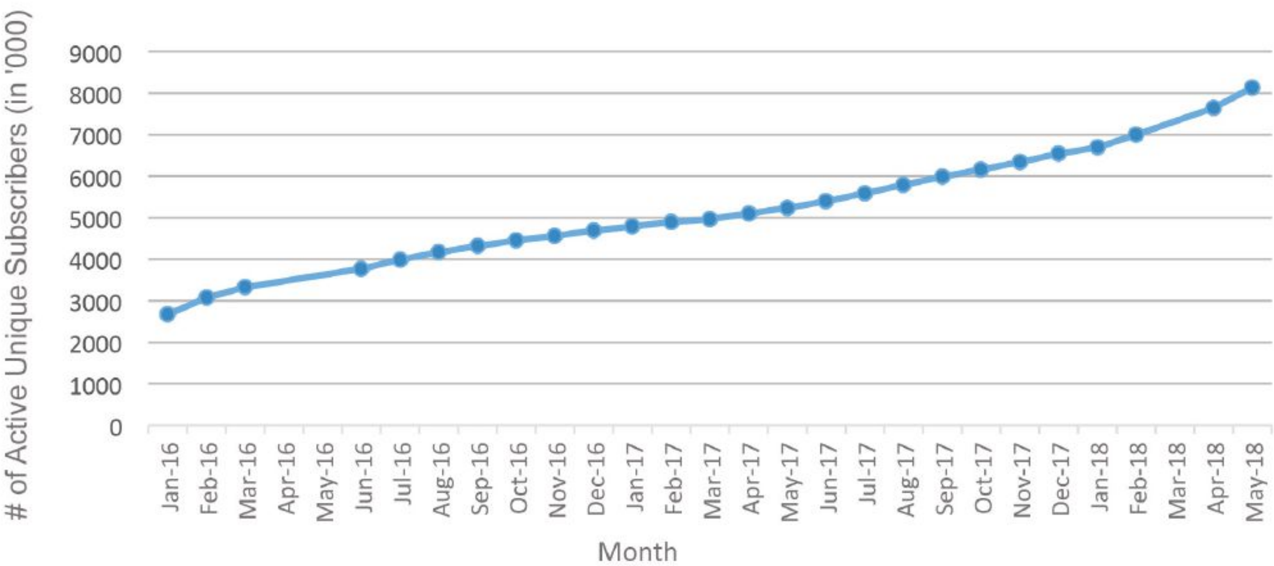
[2] WorldPop uses geospatial data calibrated on census data to create high-resolution (30 m by 30 m) maps of population distribution. See http://www.worldpop.org.uk/.
[3] The phone numbers were anonymized using a cryptographic hash function, SHA 256 algorithm. These functions are collision resistant, in that they generate a unique output, and one-way, so they cannot be decrypted back.
[4] The data is over the span of 29 months (Jan 2016-May 2018) but three months are missing (April 2016, May 2016 and March 2018)

Data Compiling and Cleaning

When reviewing the anonymized data transferred from the MNO partner, several issues were identified that required further cleaning and wrangling. Three months of data were originally missing. In addition, two months out of the 26 only had SMS data and were missing observations for calls sent and received. The analysis was trained on the whole dataset, on the assumption that these missing months were not anomalies. Trends in the growth of unique users suggest the Project Team could safely extrapolate from observed months. *See Figure 1 below.*

FIGURE 1

Rapid Increase in the Number of Unique Active MNO Subscribers



The data contained 12.9 billion records, including 23.4 million unique receiving numbers. Active unique users were defined as users having used their phones at least once in the past three months. The ratio between unique originating numbers and unique receiving numbers is high for two reasons: the originating numbers are from one MNO, while the receiving numbers could be from either of the two MNOs that are active in Malawi.[5] Furthermore, in order to minimize information loss, there were no filters applied to account for application to person (A2P) calls and messages or for the practice of "buzzing," which is a brief phone call alerting the other person of the need to call back.[6] The Project Team did not apply any filters to the records, such as A2P calls and SMS.

Therefore, there was a high chance that the ratio between the amount of unique receiving numbers to the amount of unique originating numbers would be high, especially for A2P calls. The data also contained unique location area codes (LAC IDs), and tower-level cell IDs, allowing the unique originating numbers to be matched to the nearest cell tower.[7] Furthermore, the data accounted for the number of calls and SMS separately, as well as call forwarding and roaming. *See Table 1 on page 12.*

---

[5] The two MNOs exist as a duopoly, with the one under consideration holding more than half of market share.
[6] The ratio between receiving and originating drops between 2016 and 2017, which may be tied to a government push to register all SIM cards, rendering inactive SIM cards defunct. This does not affect the analysis, which is built on the number of unique originating numbers.
[7] There were 7,025 unique cell tower IDs in the master list of towers and 6,743 unique cell IDs in the CDR data, including towers that entered and exited the dataset. There were 6,049 matching IDs across both datasets. These towers constituted the sample size in terms of coverage.

TABLE 1

Rapid Increase in the Number of Unique Active MNO Subscribers

Raw MNO Data	2016	2017	2018 (Jan-May)	Total
# Records	5,004,750,666	5,960,543,159	1,934,707,383	12,900,001,208
# Unique originating numbers	12,261,549	16,762,626	8,821,575	23,400,290
# Unique receiving numbers	50,092,409	31,288,109	18,993,028	78,600,039
# Unique LAC IDs	39	54	54	66
# Unique cell IDs	3,926	5,562	6,648	6,743
# Voice calls	3,073,839,525	3,375,532,407	592,711,535	7,042,083,467
# SMS	1,927,785,317	2,582,155,890	1,341,477,160	5,851,418,367
# Call forwarding	1,278,911	1,996,888	429,349	3,705,148
# Roaming call forwarding	1,846,913	857,974	89,339	2,794,226

The unique originating numbers were matched to the nearest tower's latitude and longitude using the cell ID. Of the 6,743 unique tower IDs in the MNO data, 6,049 were matched to the roster of cell towers with geospatial coordinates. Catchment area was calculated using Voronoi polygons, which delineate the area closest to every cell tower, with a maximum range of 20 km in rural areas.[8] Accounting for missing observations, the pre-analysis concluded that signal from the observed towers could reach a land surface where an estimated 95% of Malawi's population lives. This was considered sufficient to proceed with the analysis. The density of unique users was calculated by dividing the number of observed users by the cell tower's catchment area.

To conduct the analysis, the above MNO data was combined with several Malawi-specific data sources. This included 2015 population density at 100-meter resolution from WorldPop, calculated using satellite imagery trained on Malawi's previous census from 2009 (WorldPop 2018).

UNICEF conducted an extensive survey of every operating health facility in the country and calculated the relevant catchment area. This catchment area constituted the distance such that a patient would have to walk no more than 5 km to reach a health facility. In the analysis, UNICEF accounted for the road networks, topography, and potential for flooding in calculating the 5 km catchment area for each health facility. UNICEF calculated a "best case," where there were no impediments to travel, and a "worst case," where flooding made certain roads and health facilities inaccessible.[9] This report used the best-case model unless otherwise noted.

In collaboration with the Malawi MoH, the analysis also incorporates facility-level disease burden data, as reported on a monthly basis to the ministry.

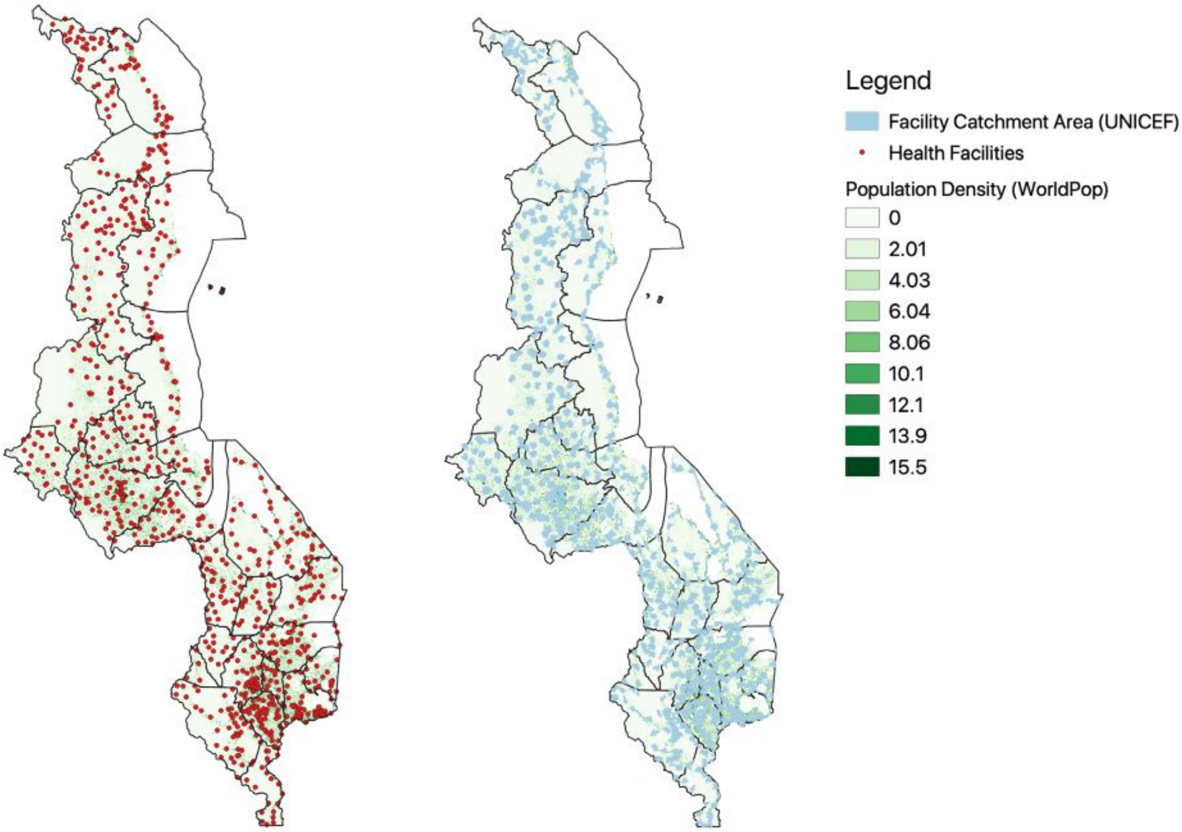
[8] Though the MNO provider shared the location of its towers, it did not provide a map of its countrywide coverage, requiring the authors to estimate this coverage instead.
[9] Special thanks to our colleagues at UNICEF for sharing these catchment area calculations with us.

GAP ANALYSIS

A first step was to establish an estimated coverage gap in terms of health services at the district and Traditional Authority (TA)[10] levels based on facility catchment areas calculated by UNICEF combined with 2015 WorldPop population data. For the purposes of this exercise, access to any facility, whether primary, secondary, or tertiary, was considered sufficient. While access to surgical and reference facilities are critical elements of a well-functioning health system, the study focused, at the government's directive, on identifying and expanding access to primary care. *See Figure 2 below.*

FIGURE 2

44.7% of Malawians Live Outside the Current Catchment Areas



Overlaying these calculated catchment areas on WorldPop population distribution made it possible to calculate the population in each district that is not within 5 km of a health facility and, therefore, lacks readily available access to primary healthcare.

The following table presents the populations with and without access in each of Malawi's districts, as both a total and a percentage of each district's population. Estimates suggest more than 450,000 people in each of five districts are currently underserved. These include Lilongwe, Mangochi, Dowa, Kasungu, and Mzimba.

[10] In Malawi, the traditional authority is the administrative unit below the district level.

Mapping the coverage gap of the TA level revealed similar patterns of coverage. Certain smaller districts have fewer people without access in total but are underserved as a percentage of the district population. For example, in Balaka only 54% of the population has access to health facilities. This suggests a trade-off between ensuring access to the maximum number of people and access to the highest percentage of people per district. *See Table 2 below.*

TABLE 2

Population with Access to Health Facilities Within 5 km

District		UNICEF Health Facility Catchment Areas Coverage		
Name	**Population**	**Total with Access**	**Total Without Access**	**% Population Without Access**
Total	**17,303,307**	**9,567,641**	**7,735,666**	**44.71%**
Balaka	421,134	227,263	193,871	46.04%
Blantyre	1,300,397	1,032,835	267,562	20.58%
Chikwawa	577,274	356,047	221,227	38.32%
Chiradzulu	381,370	273,895	107,475	28.18%
Chitipa	236,697	126,245	110,452	46.66%
Dedza	830,299	403,326	426,973	51.42%
Dowa	739,222	285,761	453,461	61.34%
Karonga	358,380	233,863	124,517	34.74%
Kasungu	829,530	267,784	561,746	67.72%
Likoma	11,962	11,962	0	0.00%
Lilongwe	2,526,221	1,565,465	960,756	38.03%
Machinga	648,531	323,784	324,747	50.07%
Mangochi	1,058,506	523,622	534,884	50.53%
Mchinji	605,201	224,441	380,760	62.91%
Mulanje	689,479	468,309	221,170	32.08%
Mwanza	122,127	50,493	71,634	58.66%
Mzimba	1,137,498	544,283	593,215	52.15%
Neno	141,353	65,379	75,974	53.75%
Nkhata Bay	286,593	126,249	160,344	55.95%
Nkhotakota	404,014	207,532	196,482	48.63%
Nsanje	314,478	211,262	103,216	32.82%
Ntcheu	623,126	317,567	305,559	49.04%
Ntchisi	298,223	126,582	171,641	57.55%
Phalombe	416,471	212,139	204,332	49.06%
Rumphi	232,241	126,774	105,467	45.41%
Salima	448,545	219,501	229,044	51.06%
Thyolo	777,455	473,007	304,448	39.16%
Zomba	886,982	562,268	324,714	36.61%

POPULATION MOVEMENT DYNAMICS

Estimating Population Density

Having identified a gap in coverage based on static population patterns, the next step was to incorporate MNO data analytics to account for population growth and migration patterns.

For each cell tower, the Project Team identified the number of unique users within the cell tower's catchment area.[11] Dividing the number of unique users by the tower's catchment area, revealed the density per tower. This density was mapped to TAs by calculating the weighted average of overlapping polygons, providing the density of unique originating IDs in each TA.

The analysis sought to determine the relationship between the density of unique originating IDs σc and population density P^c. The training data for population density came from the 2015 WorldPop data, extrapolated to 2016 and 2017, and calculated for each TA to ensure one-to-one correspondence between σ^c and P^c.

The relationship was estimated using a linear regression formula estimated using ordinary least squares (OLS):

$$logP^c = \alpha + \beta log\sigma^c + \mu_k$$

Where α is the regression constant, β is the coefficient of interest, and μ_k is a regional fixed effect that allows for inter-regional variation. The regression was calculated separately for 2016 and 2017. To allow for potential spatial correlation, the specification was tested for the significance of Moran's I, a matrix weighted for spatial adjacency (Odland 1988). If the null of no spatial correlation was rejected, the model was estimated to account for spatial co-variance.

The central region of the country has a higher beta coefficient than the south or the north, likely because it has a much higher mobile penetration, as evidenced by the ratio of population density to unique call density. This validated the need to separate coefficients across regions. The coefficients were consistent within regions between years. *See Table 3 on page 16.*

[11] Unique users were defined as those active within the last three months (at least one SMS or call). As in Deville et al. (2014), the number of unique users per tower was calculated as the sum of unique users active within the cell tower's polygon between 8:00 pm and 7:00 am, when users are assumed to be at home. Their location was determined as the tower they use most frequently, the model tower. The number of unique users per night was then averaged over the entire year.

TABLE 3

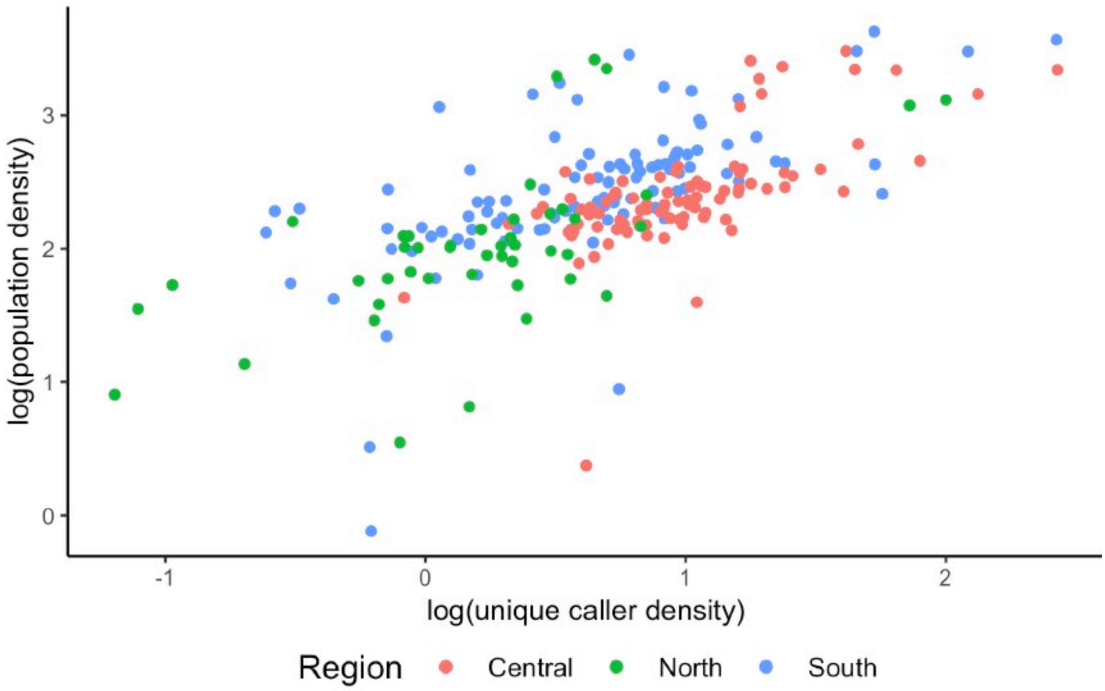
Results of Regressing Population Density on Density of Unique Callers, OLS

	2016			2017		
	South	Central	North	South	Central	North
α	4.52	3.43	4.1	4.51	3.29	4.01
β	0.624	0.86	0.595	0.597	0.89	0.597
P-Value	(6.815*10^-9)	(0.00478)	(0.0002662)	(3.388*10^-8)	(0.0349)	(0.0002321)
R^2	0.2748	0.3701	0.2533	0.2529	0.3707	0.2575
Moran's I	0.108	0.021	0.344	0.197	0.026	0.36822
Account for Spatial Correllation	No	Yes	No	No	Yes	No
(P^c/σ^c)	60.83	19.81	59.64	51.56	17.41	49.92
N	107	95	48	107	95	48

Spatial k-fold cross-validation was used to validate the above results (Pohjankukka 2017). The model was re-run k=8 times, each time omitting 1/8 of the sample. The estimated coefficient $\hat{\beta}$ was then used to calculate the predicted population density *logP*^{c} and compared to the actual population density. The R^2 values suggested that variation in unique caller density accounts for between a quarter and a third of the variation in population density. *See Figure 3 below.*

FIGURE 3

Correspondence Between Log Call Density and Log Population Density



As evidenced from the figure above, the relationship between log call density and log population density was largely linear; though, there were some outliers at the low-end of population density, including sparsely inhabited areas where the number of unique callers is an imperfect proxy for population.

Unlike the static population estimates from the WorldPop data, the number of unique user IDs from the MNO data varied from day to day and month to month. As seen above, the regional coefficients were consistent between 2016 and 2017, tentative evidence that the relationship holds over time. The estimated coefficients can therefore be used to calculate predicted population densities across time, uncovering the dynamics of population movement.

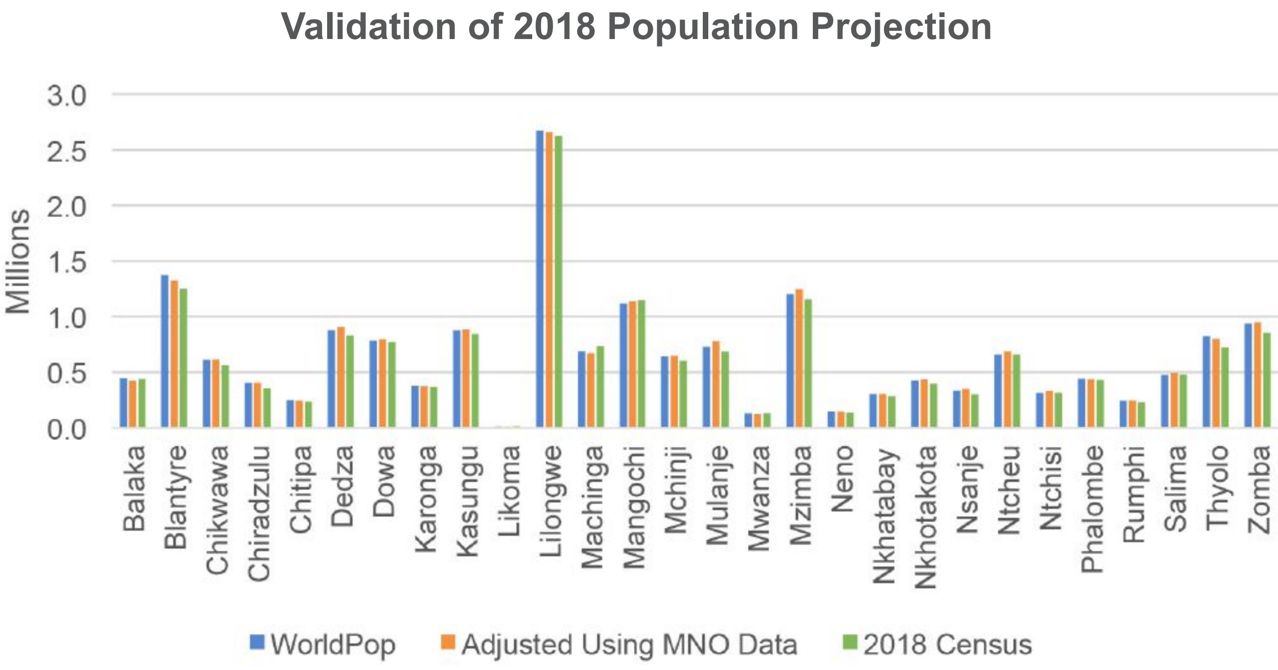
Validation Using Census Data

With the beta coefficients estimated, the next step was to project population growth patterns forward for every TA. The static WorldPop data used a uniform growth rate across all TAs to project population growth, assuming that it will grow at a rate of 2.785% per year across all administrative areas.[12] Instead, the Project Team calculated the change in the density of unique users across time in each TA. In order to account for subscriber growth, a unique set of active users were identified at the beginning of January 2016 and their movements tracked throughout the two years to capture shifts in population.

Using the coefficients estimated above, these shifts in unique users were used to estimate the percentage change in population density for each TA. While the overall country's population growth rate was held constant, each TA grew at a different rate. Certain TA's, largely in urban areas, were allowed to grow faster, while other TAs, notably in rural areas, grew slower or even shrunk. The country's population was therefore redistributed to allow for observed migratory patterns.

In order to validate this methodology, the predicted population levels were compared with the results from the 2018 census. The census provides a detailed headcount of population based on on-the-ground surveys at the national, district, and TA levels in 2018. This is considered the authoritative, ground-truth data. This was compared with predicted population levels based on: (a) the WorldPop data projected uniformly forward to 2018[13] and (b) the TA-specific growth rates inferred from shifts in a fixed number of unique users. So, while the totals for (a) and (b) were identical, the district level population levels differ. *See Figure 4 below.*

FIGURE 4



There was less than 5% discrepancy between the predicted and actual population levels for each district, revealing that—in Malawi—satellite data and MNO data analytics were both good proxies for actual population levels.[14]

[12] This number was derived by comparing TA level population levels as projected by WorldPop in 2015, 2016 and 2017.
[13] Recall that WorldPop based its projections on the 2009 census, training satellite data on it, and projecting forward using a uniform growth rate.
[14] While this validates long-term growth patterns, the Project Team do not know the composition of cell phone users relative to those who tend to migrate, short- and medium-term population, so mobility may be overstated or understated.

Short- and Medium-Term Population Movements

Once benchmarked to population density, the Project Team calculated population movement from month-to-month and across times of day. This allowed for the following three types of analysis:

1) Commuting

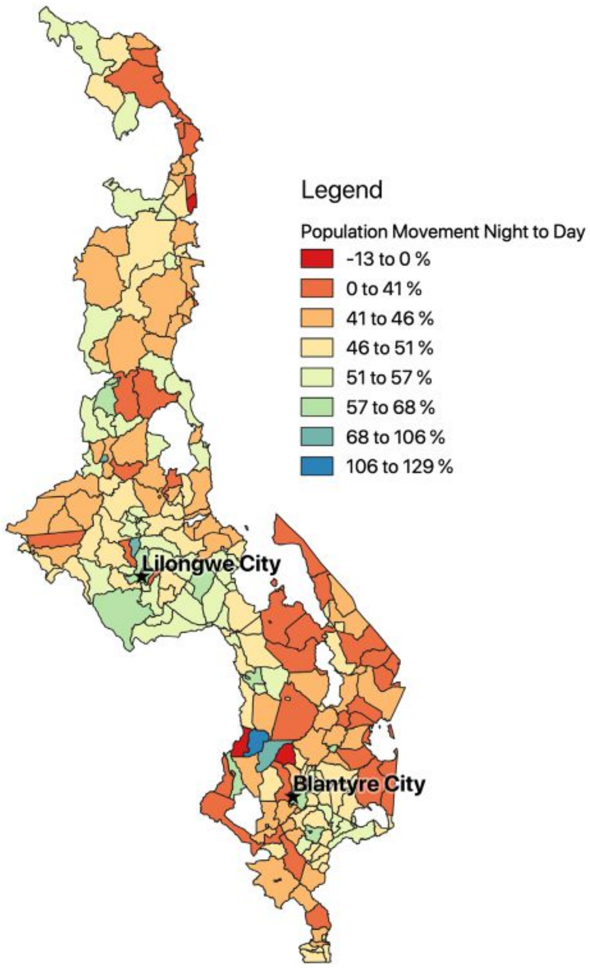
To understand commuting patterns, the Project Team compared the population as inferred by cell phone activity during the nighttime (8:00 pm – 7:00 am) and daytime (7:00 am – 8:00 pm).[15] The map below presents the percentage change in estimated population density during the day relative to the night. The percentage changes are mostly >0, reflecting a higher level of cell activity during the day relative to the night.

Nighttime location was assumed to indicate where a person lives. Shifts in their movement during the daytime offer evidence of commuting behavior. This may be relevant for patients seeking care when going to or returning from work.

Comparing nighttime and daytime activity at the TA level, there was a general surge during the daytime and some evidence of commuting into the larger cities. The report accounted for these commuting patterns in recommending the location of health posts by emphasizing areas where people spend the night, between 8pm and 7am. *See Figure 5 below.*

FIGURE 5

Evidence of Commuting from Periphery to City Centers



[15] Each unique number's daytime and nighttime location was inferred using their modal location during the daytime and nighttime.

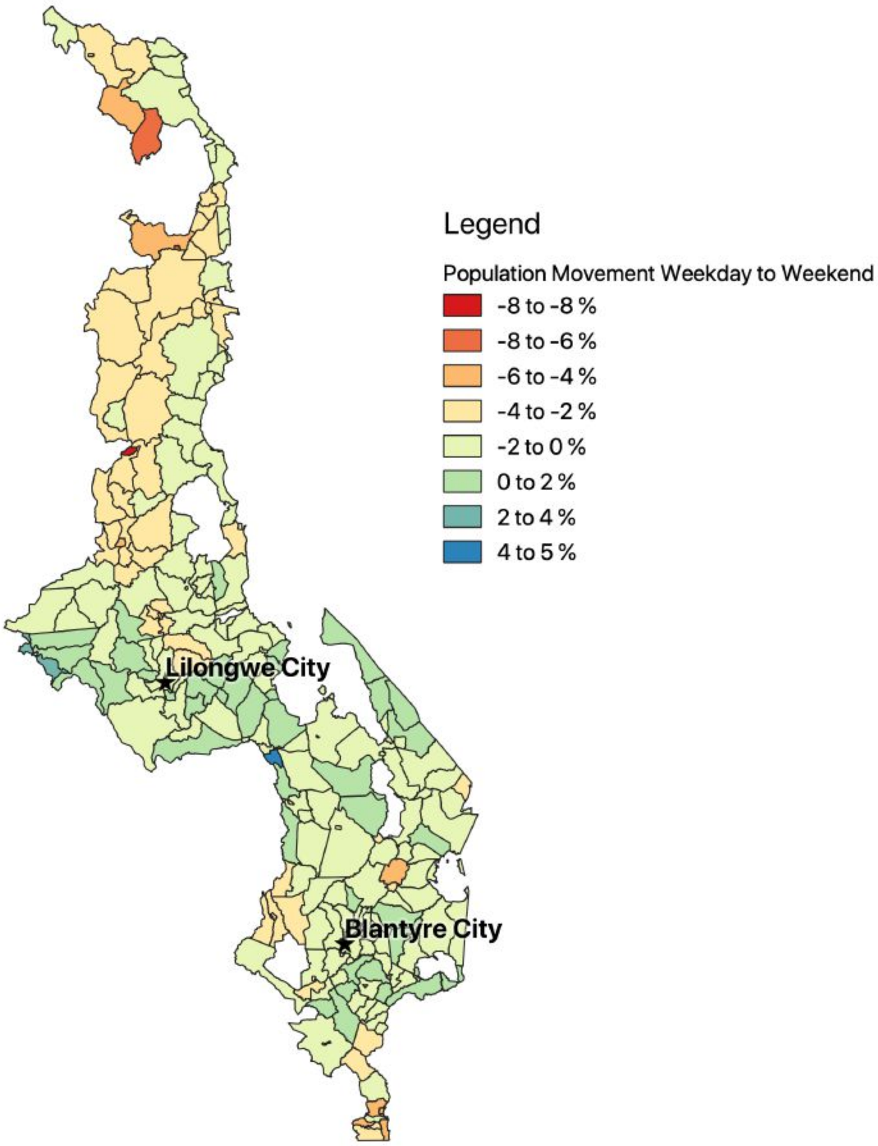
2) Weekend vs. Weekday

A second analysis observed population movement across the weekday and weekend to infer whether there were significant population movements to be accounted for when providing health services. This was done by comparing the average number of nighttime unique users per TA during the week to the average nighttime unique users per TA during the weekend and extrapolating to population using the estimated coefficients. The map below displays the percentage change in estimated population density on weekends relative to weekdays.

The most significant shift seemed to be in the north when, on weekends, a large number of people move from west to east, potentially to the lake. In the center, there is some evidence that individuals leave Lilongwe on weekends and that there are numbers of people along the Mozambique border over the weekend, suggesting the presence of marketplaces.[16] *See Figure 6 below.*

FIGURE 6

Population Shifts to Coast and Markets on Weekends



[16] This is an observation derived from the work the Project Team was doing alongside the government in the region.

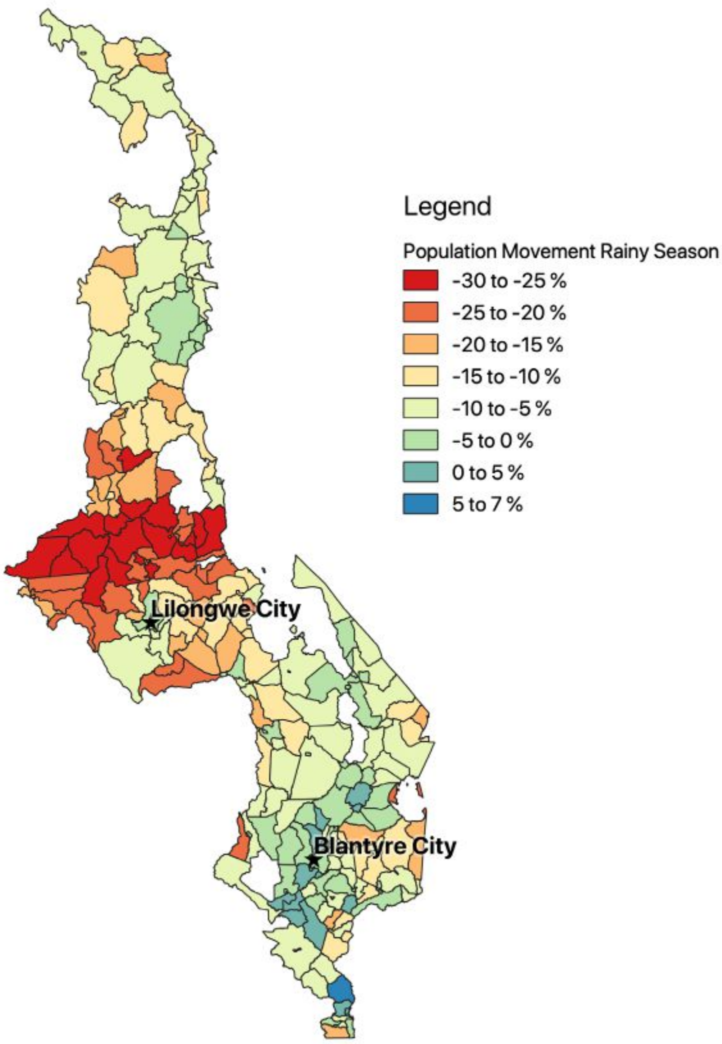
3) Migration

A third area of analysis was seasonal migration. Since an estimated 80% of Malawi's population engage in agricultural activities, many of them are subject to seasonal migration, particularly seasonal laborers. The Project Team therefore compared population density in the rainy season (November - April) to the non-rainy season (May - October). This meant comparing the average number of active unique users per month[17] in the rainy and dry seasons and extrapolating to the population level using the coefficients outlined above.[18]

The map below presents the change in population density, illustrating population movement during the rainy season as a percent change. *See Figure 7 below.*

FIGURE 7

Population Shifts from the Central Region to the Southern Region During the Rainy Season



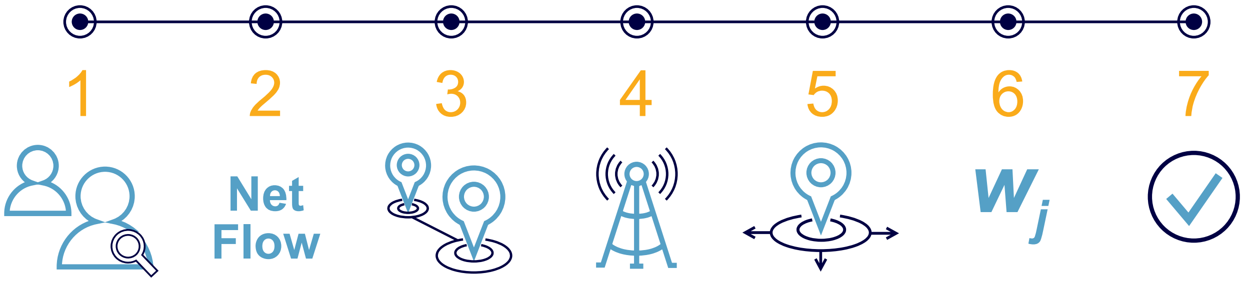
During the rainy season there is a large-scale migration from the center of Malawi, especially around Lilongwe, towards the south. In particular, populations shift towards the Shire River Basin, an agriculturally fertile land where intensive agriculture is practiced. This offers evidence of seasonal migration driven by the need for labor for agricultural activities.

---

[17] This includes all users who sent an SMS or made a call in the past three months.
[18] As the demographic characteristics of cell phone users are not available, these estimates may either be an upper bound, if those who migrate are more likely to own cell phones, or a lower bound, if those who migrate are poorer and thus less likely to own cell phones. The magnitude of the shift suggests an upper bound estimate.

OPTIMIZATION OF FACILITY ALLOCATION

Given the observed gap in service provision, an optimal allocation of health posts can be calculated using projected population growth based on MNO data analytics. This optimization seeks to maximize coverage of currently unserved populations, accounting for population growth and migration patterns.



The analysis proceeded in the following sequence:

STEP 1

Identified a consistent set of unique users in 2016 and 2017 and **calculated the net-flows in or out of each Traditional Authority** for each year. The net-flow is the difference between number of unique users in the TA at the beginning of the year and number of unique users in the TA at the end of the year. Restricting the analysis to a consistent sub-sample ensures that additional subscribers do not bias the sample. The TA-level 2016 and 2017 net-flows were not statistically different, so the average was taken.

STEP 2

Used NetFlow to adjust projected population growth for each TA. As described above, the national-level projected growth rate was pegged to the WorldPop growth rate, but TA level growth was allowed to vary. This method was validated using the 2018 census. Some TAs grew faster, especially in urban areas, and some TAs grew slower or even shrank. Population growth was then projected forward to 2019 to 2023, allowing differential growth rates that more accurately reflect migration and urbanization patterns.

STEP 3

Used the catchment areas of existing health facilities defined by UNICEF, along with **population density** and **long-term population movement**, to identify the availability of health facilities within each TA relative to the population. The total estimated uncovered population at time t is U_t.

STEP 4

Calculated spatial clusters containing a given uncovered population based on the WorldPop data adjusted for population growth and migration. To obtain such clusters, population density was calculated for each cell tower catchment area with the highest resolution available, using the estimated coefficients and adjusting for TA-level population growth. Short-term movements were used to identify highest-used locations within each TA. These clusters were restricted to be outside of the existing catchment areas, have a maximum radius of 6 km and contain no more than 12,500 people. If a given area contained more than 12,500 people, it was split, and two new clusters were calculated. The centroid for these spatial clusters was identified. These demand points are the candidate sites for new facilities, denoted P_{ij} for facility i in TA j.

STEP 5

Selected a distance beyond which a patient cannot travel to the health post: D called Impedance cutoff. This defines the catchment area. To maximize coverage, the cutoff was allowed to vary between 5 and 6 km. The population within this catchment is denoted C_i.

STEP 6

At the TA level, the **importance of population can be mediated by weight** w_j capturing TA-level disease burdens.

STEP 7

Optimization was conducted iteratively for every year, accounting for health posts allocated to date. Given the demand points, the number of facilities allocated in a given year is optimized to minimize the number of under-served U_t given the constraint $\bar{P}$, the maximum number of health centers that can be built within a year. Once allocated, in the subsequent year the demand point P_{ij} is removed from consideration, since it's already allocated. The model seeks to optimize the following objective function:

$$\text{Min } U_t - \sum\sum w_j * \sum P_{ij} * C_i$$

$$s.t \sum P_{ij} \leq \bar{P}$$

Based on the above, three different models were run.

MODEL 1

Accounts for underserved population only, giving all TAs an equal weight w_j of 1.

MODEL 2

Accounts for underserved population and TA-level disease burden but doesn't rank the severity of disease. w_j is the number of patients at the TA level, regardless of disease.

MODEL 3

Accounts for underserved population, TA-level disease burden and the severity of disease, ranking these by severity in terms of disability adjusted life years (DALYs) from the Global Burden of Disease (GBD) compiled by the Institute for Health Metrics and Evaluation (IHME).[19] w_j is the number of patients at the TA level, weighted by disease.

---

[19] Available at http://www.healthdata.org/malawi.

Model 1: Underserved Population Only

The results found that by strategically placing 900 new health posts to account for existing gaps, 95% of the population would be within 5 to 6 km of a health post by 2023. Adjusting for disease burden does not significantly alter these results. If those health posts are not built, 9.7 million Malawians, 44.85% of the population, would still be uncovered by 2023. *See Table 4 below.*

TABLE 4

Results from Model 1, 2 and 3 in Terms of Coverage

Descriptions	Model 1	Model 2	Model 3
Year 4 (2023) forecasted population	21,621,892	21,621,892	21,621,892
Year 4 (2023) estimated uncovered population before 900 new health posts	9,696,818	9,696,818	9,696,818
Year 4 (2023) estimated uncovered % population before 900 new health posts	44.85%	44.85%	44.85%
Year 4 (2023) estimated uncovered population after 900 new health posts	1,122,720	1,228,322	1,242,047
900 new health posts	5.19%	5.68%	5.74%

These health posts are each expected to serve a maximum of between 12,000 and 12,500 people within a 5 to 6 km radius. The proposed schedule of construction is drawn from Malawi's Capital Investment Plan, with the assumption that one health post scheduled for 2019 will be built in 2020 instead, for a total of 198 health posts in 2020. In each subsequent year, 234 health posts are to be built in order to reach the goal of 900 health posts by 2023. *See Table 5 below.*

TABLE 5

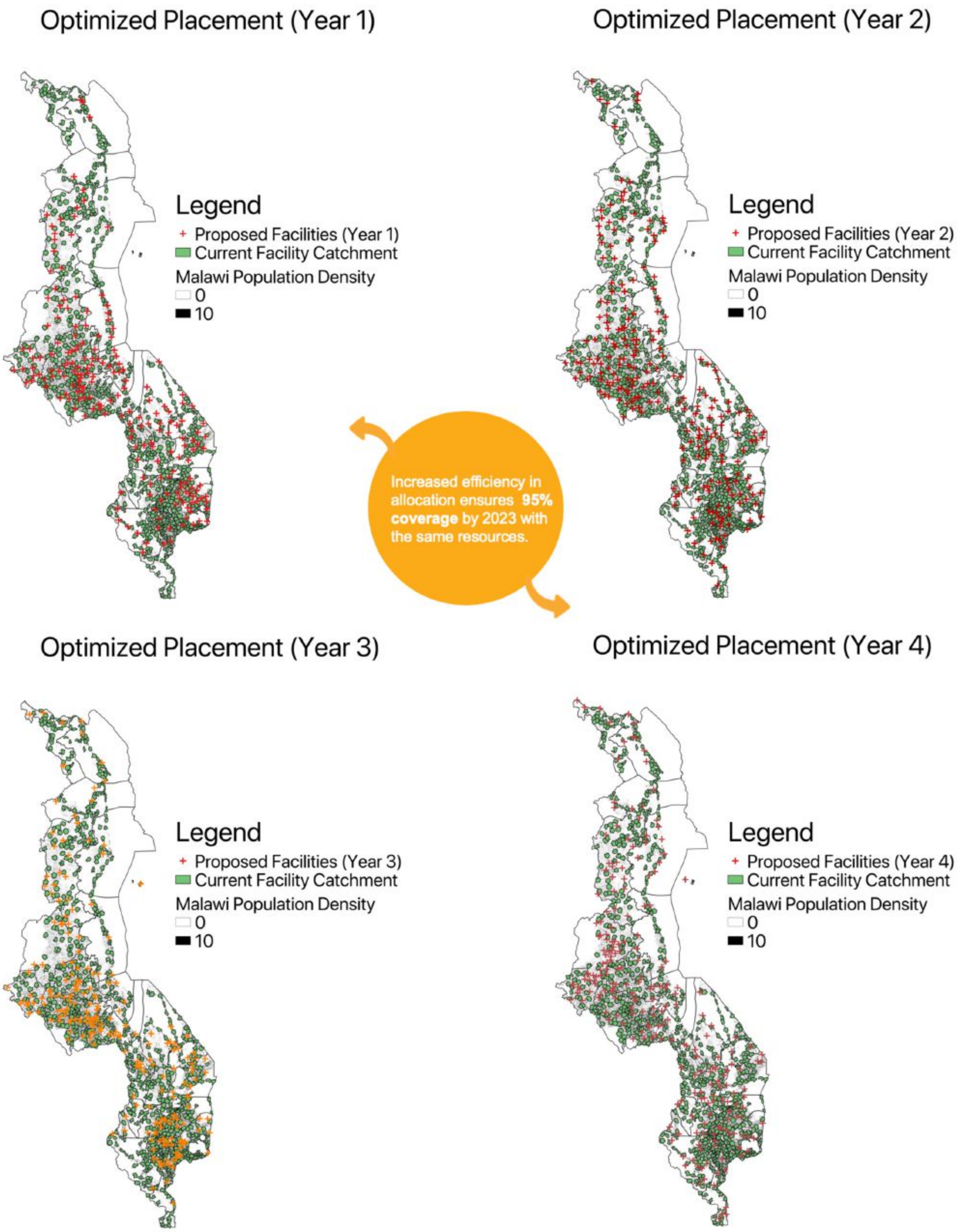
Year-on-Year Distribution of Health Posts by Catchment Population (Model 1)

Year	Up to 12,000	12,000 to 12,500	Total
1	108	90	198
2	181	53	234
3	209	25	234
4	233	1	234
Total	**731**	**169**	**900**

The allocation of these new health posts reflects population growth patterns, with the model seeking to fill gaps in coverage in both rural areas and rapidly expanding urban areas. The initial facilities are expected to serve upwards of 12,000 people each, reflecting the pent-up demand for services. As new facilities are built, the additional number of people served by each health post gradually goes down. The 900th health post is expected to serve fewer than 3,000 people living within 6 km. This model is currently informed by historical MNO data but can be updated using periodic or close to real-time MNO data as it becomes available. *See Figure 8 on page 24.*

FIGURE 8

Proposed Allocation of New Health Posts



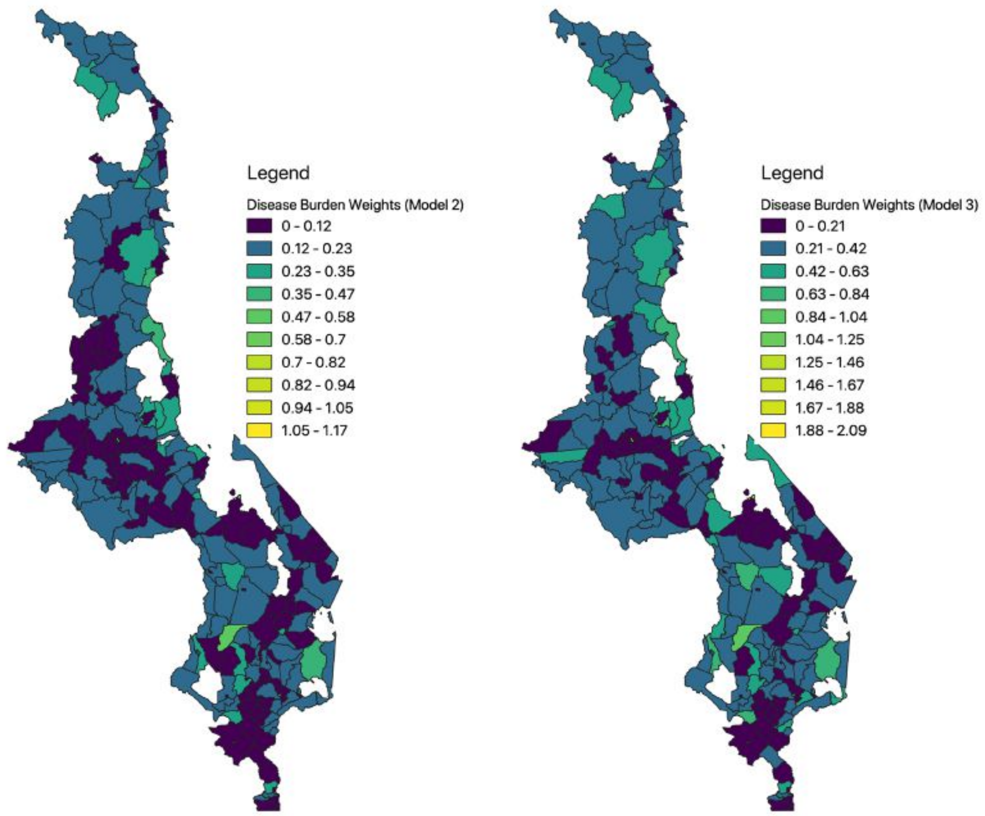
Model 2: Underserved Population and Disease Burden

Model 2 incorporates TA-level disease burden as reported by the Ministry of Health through the Health Management Information System. These were reported as 34 indicators, reflecting the number of patients diagnosed and receiving treatments every month in each reporting facility. The Project Team chose to use administrative data because of both its timeliness and granularity relative to survey data, while allowing for the fact that cases reported may not necessarily reflect the actual disease burden in all areas due to suppressed demand.

In order to create weights from this HMIS data, each indicator was first summed across the year to reflect the total annual disease burden. Indicators from each facility were then summed up to TA-level. To accurately compare disease burdens across TAs, they were normalized by dividing the sum or reported cases by the TA population. An exception was made for district and central hospitals, which were assumed to service the entire district rather than a single TA. Their disease burden was therefore allocated to each TA in the district in proportion to the population, rather than to a single TA. *See Figure 9 below.*

FIGURE 9

Disease Burden Weights Based on HMIS Data



Finally, to calculate the weights, the 34 indicators were summed into a single index to reflect overall disease burden. This raw sum included the number of patients, cases, and deaths across all categories. As the administrative data does not differentiate by patient, some of these indicators, such as OPD visits, included potential repeat visits by the same patient for the same condition. While an imperfect approximation for disease burden, it does reflect the workload health facilities currently face.

In terms of results, TA-level disease burdens were incorporated as weights to the objective function outlined at the top of page 18, and therefore skewed the allocation of facilities towards TAs that had a higher burden of disease. These weights adjusted the prioritization of allocation slightly, though the overall coverage after four years was near identical at an estimated 94.32% of the population.

Model 3: Underserved Population, Disease Burden and DALY

Model 3 differs from model 2 in that, in the final step, the sum of HMIS indicators incorporated the severity of the disease in terms of disability adjusted life years (DALYs), as determined by the Global Burden of Disease compiled by the Institute for Health Metrics and Evaluation (IHME). The weights were on a scale of 1 to 5, with 5 being the most severe. Adjusting for DALYs therefore put more weight on the indicators that have the greatest impact on health. *See Table 6 below.*

TABLE 6

Disease	Weight Without DALY	Weight With DALY
# of 15 - 49 age group tested HIV positive	1	5
# of HIV positive persons receiving ARV treatment	1	1
# of HIV positive women treated for PMTCT	1	5
acute respiratory infections - new cases (u5)	1	5
HIV confirmed positive (15-19 years) new case	1	5
diarrhea non - bloody -new cases (under5)	1	4
malaria - inpatient deaths under 5	1	4
malaria -inpatient deaths (5 & over)	1	4
malaria new case (under 5)	1	4
malaria- new cases (5 & above)	1	4
malnutrition -inpatient deaths (under 5)	1	4
malnutrition new case (under 5)	1	4
# of deliveries attended by skilled health personnel	1	3
# of direct obstetric deaths in facility	1	3
# of postpartum care within 2 weeks of delivery	1	3
# of pregnant women starting antenatal care	1	3
# of road accidents - inpatient deaths	1	3
cholera - inpatients deaths	1	3
dysentery- inpatients deaths	1	3
total # of live births	1	3
# of fully immunized under 1 child	1	2
# of persons receiving Depo-Provera	1	2
# of persons receiving IUCD	1	2
# of persons receiving Norplant	1	2
# of persons receiving 3 months' supply of condoms	1	2
# of under 1 children given BCG	1	2
# of under 1 children given pentavalent	1	2
cholera - confirmed new cases	1	2
dysentery - new cases	1	2
ear infections - new cases	1	2
measles - confirmed new cases	1	2
# of OPD attendance	1	1
total # of discharges	1	1

Results in terms of allocation were very similar to model 2, with an ANOVA test rejecting significant differences between all three models at the district and TA levels. Model 2 and Model 3 tended to produce more balanced allocations in terms of catchment populations.

ALIGNMENT WITH THE MOH CAPITAL INVESTMENT PLAN

Overview of CIP

The Malawi Ministry of Health is planning to roll out 900 health posts over the next five years across all 28 districts of Malawi. These include both upgrades to existing facilities and the construction of new buildings to expand access, particularly in rural and remote areas, with an emphasis on the provision of primary health care.

The Malawi Ministry of Health is planning to roll out 900 health posts over the next five years across all 28 districts of Malawi.

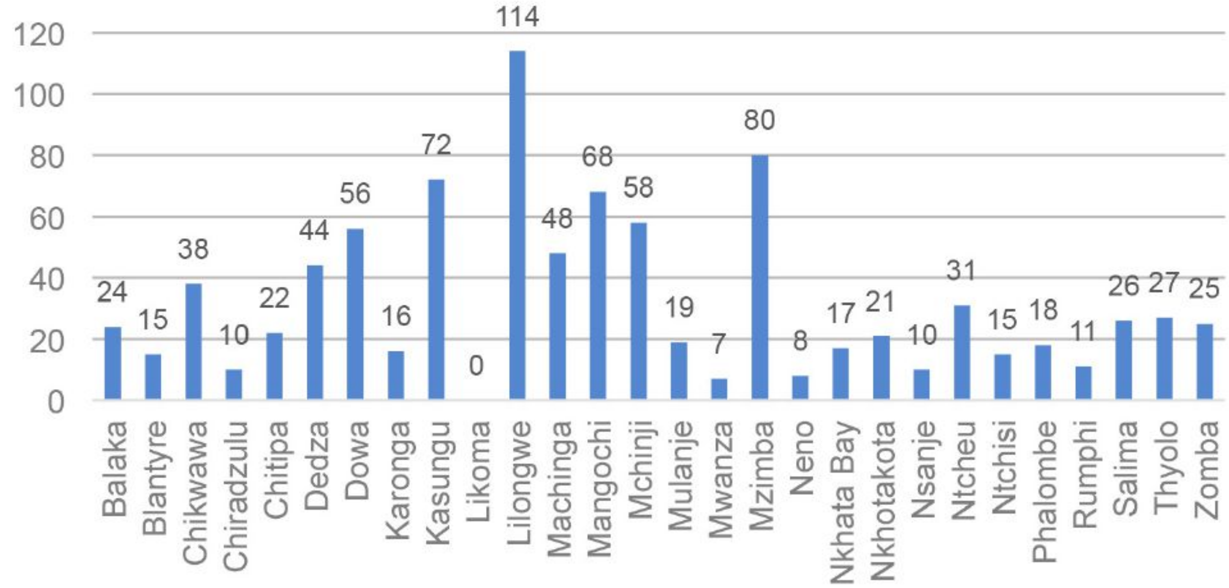
As part of its draft Capital Investment Plan, the government of Malawi developed proposed allocations of new facilities using the following four criteria:

1) Catchment population
2) Distance to nearest existing health facilities
3) Facility accessibility (high, medium, low)
4) Preferred year for work to take place as expressed by the District Health Monitoring Teams

The CIP projects that the cost of equipping and building these health posts will be $41,954,407 over five years. Based on a review of the CIP appendix, the plan recommends the allocation of health posts per district listed below in Figure 10.

FIGURE 10

Proposed New Health Posts in Capital Investment Plan



Allocative Efficiency

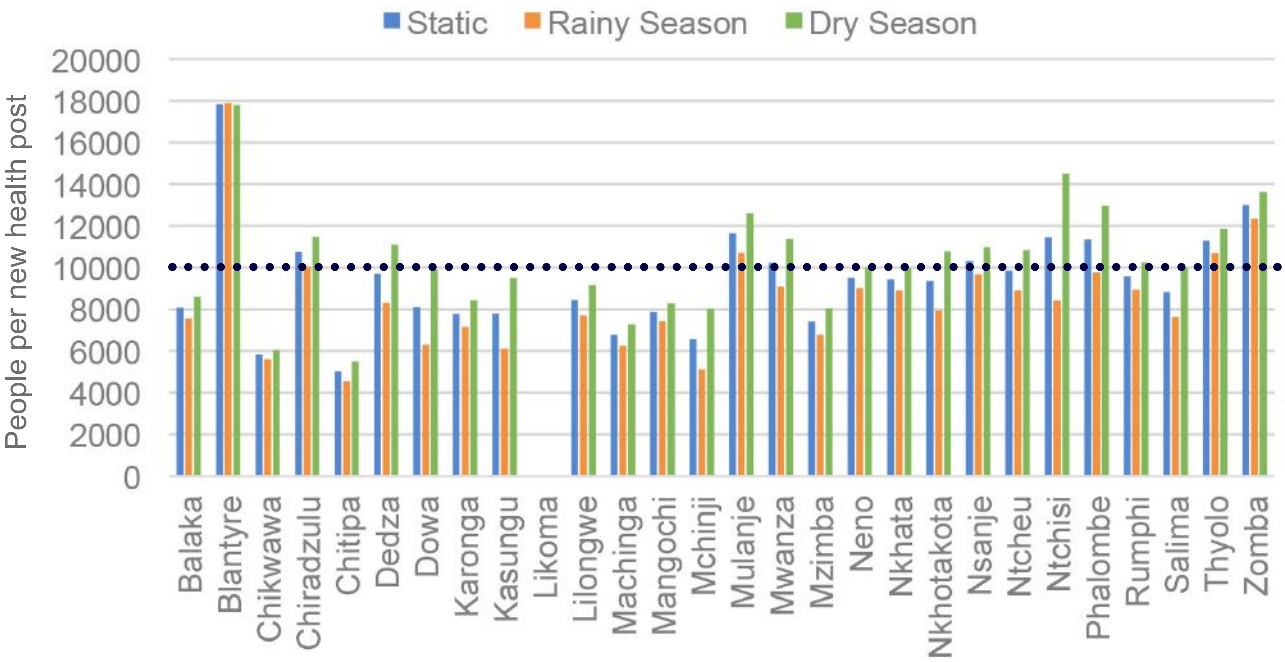
Since the exact location of these proposed health posts is yet to be determined, the Project Team could not conduct an analysis of their proposed catchment area. Instead, by incorporating the UNICEFF catchment data into the analysis presented above, it could calculate the "people per health post." That is, based on the current number of people without access per district, how many people each new health post would service assuming they service everyone without current access. People per health post therefore measures allocative efficiency.

$$\frac{\text{(\# of People Without Access in District)}}{\text{(\# of New Health Posts)}} = \text{People per Health Post}$$

Incorporating MNO data, the metric changed when accounting for population shifts in the rainy and dry seasons. The people per health post for each scenario is presented below in Figure 11.

FIGURE 11

Efficiency: Allocation of People per Health Post Under CIP



The dotted line represents an illustrative threshold of 10,000 per new health post, showing that certain districts are below that threshold and others above it.[20] Blantyre has a particularly high people per health post value, suggesting there are insufficient proposed posts to provide services to the population currently not covered.

[20] This threshold is illustrative only, as the CIP does not specify the estimated maximum capacity of a health post. The Project Team recommend consulting with MoH to establish a recommended threshold.

Allocative efficiency can also be adjusted to allow for catastrophic shocks, such as floods cutting off access to health facilities. By combining the UNICEF worst case scenario for flooding with the population movement analytics, the people per new health post ratio can be calculated in the event of catastrophic floods (such as the floods in 2015) and thereby identify the most vulnerable districts. *See Figure 12 below.*

FIGURE 12

Resilience: People per Health Post, Accounting for Flooding

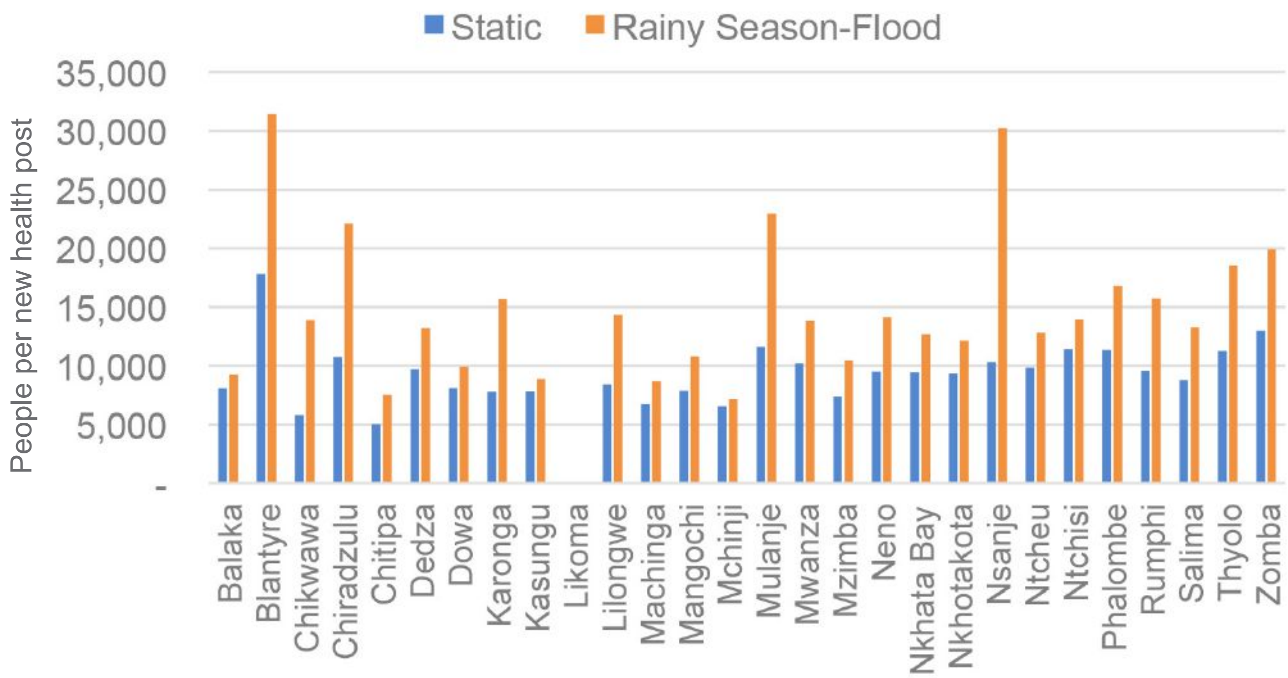


Figure 12 shows that Blantyre, Nsanje, Chiradzulu, and Mulanje are particularly vulnerable in the event of a flood. In Blantyre and Nsanje, a flood may mean that each new health post would have to service up to 31,500 people, straining its capacity.

COMPARISON WITH OPTIMIZED ALLOCATION

Comparison between Malawi's Capital Investment Plan and optimal allocation models could only be made at the district level, as the plan does not recommend TA-level allocations. The optimization model, therefore, added value by providing additional resolution and recommending both the TA and the community where new health posts should be built.

Based on the model, if 900 health posts are built in optimal locations, 95 percent of Malawians will live within walking distance of a health post by 2023.

At the district level, though the allocations were broadly aligned, there were significant differences between the allocation of new health posts as outlined in the Capital Investment Plan and those determined by the allocation. *See Figure 13 on page 31*. These differed most in Blantyre, Zomba, and Chikwawa, reflecting the difference in terms of allocative efficiency. In Blantyre alone, the re-allocation would reduce the people per health post from 17,800 to 9,200.

As a counterfactual, if each district built the number of health posts recommended in the optimized model and each health post could serve no more than 12,500 people, an additional 226,000 Malawians would have access to health services, relative to the allocation under the Capital Investment Plan.[21] *See Figure 14 on page 31.*

Given that these regions are most vulnerable to having health facilities cut off in times of flood, building additional facilities may also increase their resilience, ensuring that Malawians have access to health facilities when they need them most. In the worst flood-prone districts, the maximum number of people per health post drops by between one-third and one-half, and the highest burden drops from 31,500 to 23,200, a more manageable number in times of crisis. *See Figure 15 on page 31.*

Based on the model, if 900 health posts are built in optimal locations, 95% of Malawians will live within walking distance of a health post by 2023.

[21] This was calculated as $\sum[(U_k - P_k^{CIP} * 12{,}500) * 1(U_k - P_k^{CIP} * 12{,}500 > 0)]$ the sum of the difference between U_k, the unserved population per district adjusted for population growth, and P_k^{CIP} the number of health posts proposed by the CIP multiplied by 12,500, but only if the difference is positive.

FIGURE 13

Change in Allocation Relative to the CIP

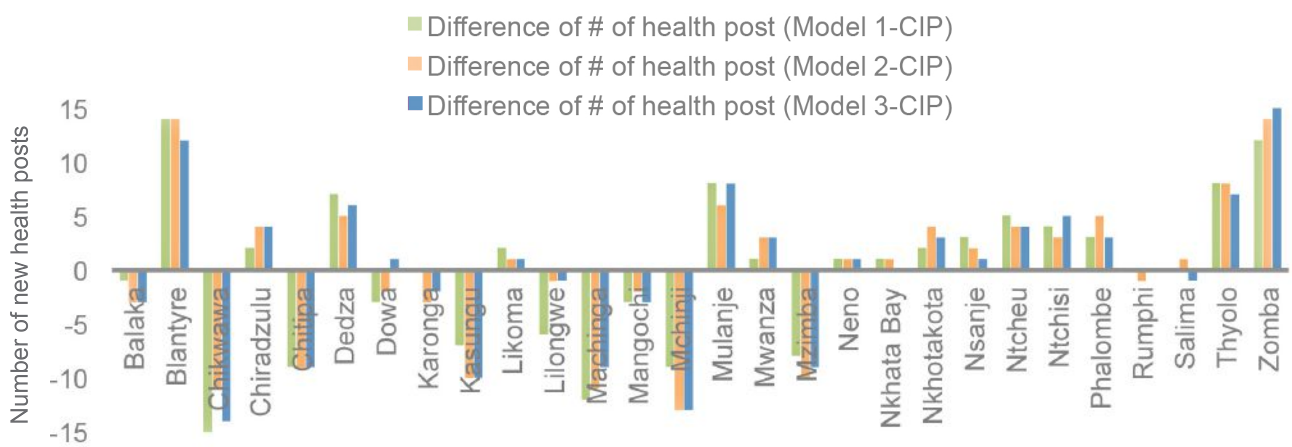


FIGURE 14

Counterfactuals: Optimized Model Improves the Efficiency of Allocation…

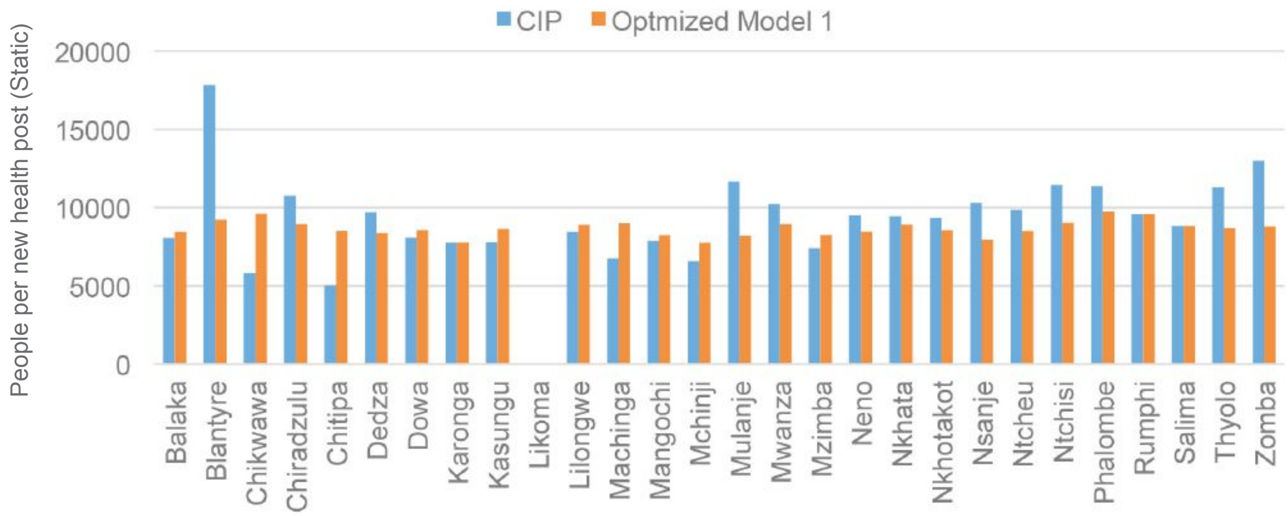
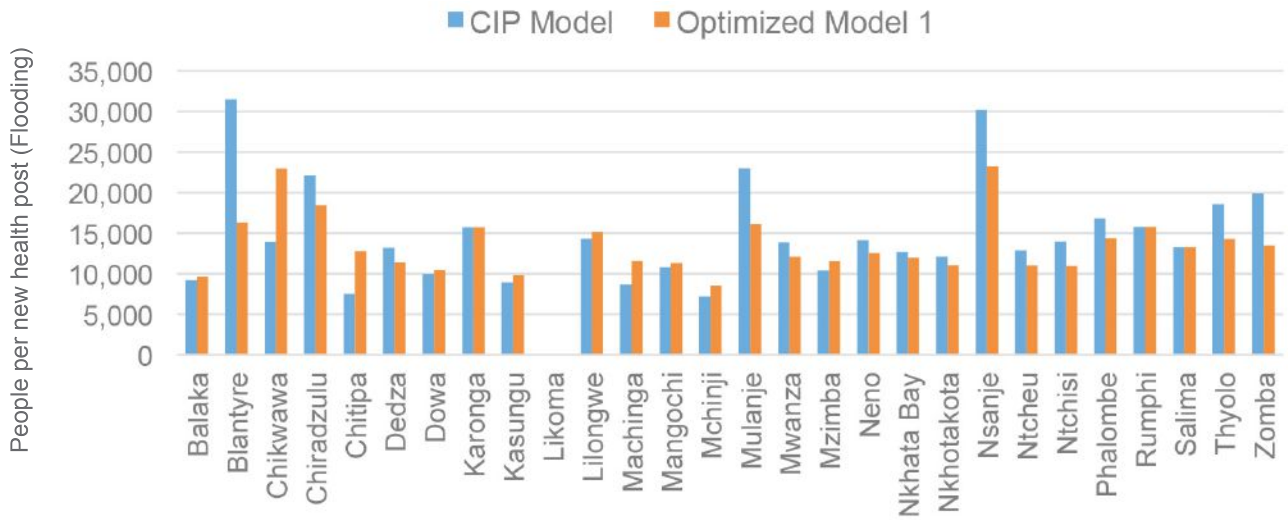


FIGURE 15

… and Its Resilience to Flooding



SUSTAINABILITY AND SUPPORT TO GOVERNMENT PARTNERS

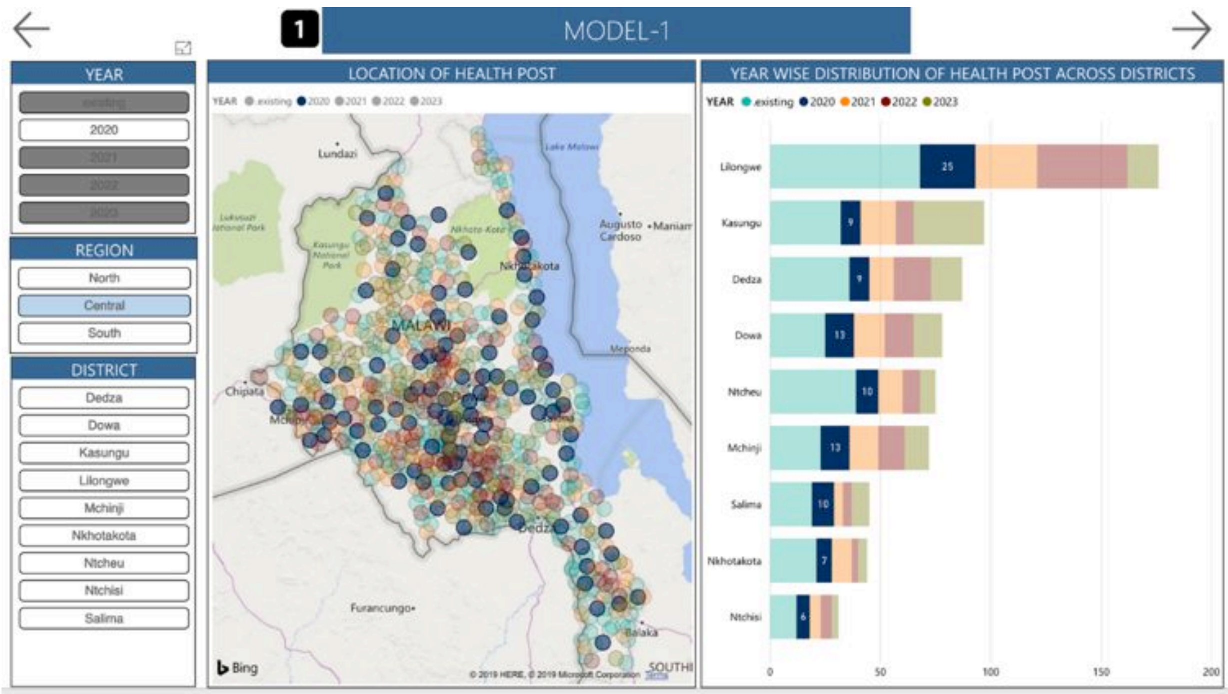
Dashboard

To provide further insights, the Project Team developed an interactive dashboard using Power BI. This dashboard provides an overview of:

- Estimated population density
- Health post coverage
- Cell phone usage patterns
- Long-term population movements
- Short-term population movements

FIGURE 16

Snapshot of User Dashboard



The Project Team intend to integrate data on patients and disease burden into the dashboard, combining these with the above in a user-friendly format based on feedback from the MoH.

Engagement with Ministry of Health

The key to ensuring sustainability, replicability, and relevance of this use case was to engage with government counterparts at both the policy and technical levels. Input was solicited throughout the process and incorporated into the model. Direct engagement around the analytical products allowed technical counterparts to provide feedback. These inputs guided the Project Team in designing analytical products around the requirements and capacities of the ministry.

In order to ensure sustainability, DIAL and Cooper/Smith organized a series of dissemination exercises. These included one-on-one meetings with government counterparts, deep-dive presentations where input was solicited, and a dissemination workshop with a detailed walk-through of the analytical products. Participants came away with an understanding of how the products inform their activities and provide concrete feedback the Project Team could incorporate to ensure the products fit into existing country systems. This process also initiated the conversation around sustainability, in order to integrate the use case into the ministry's decision-making.

These conversations around sustainability emphasized tapping into a broader set of use cases to answer other questions regarding the provision of health services. This included conversations around additional datasets that could be combined with MNO data analytics. The conversations also emphasized the importance of using appropriate tools that ministry counterparts were familiar with. In order to turn this case study into a viable long-term solution, the one-time transfer of deidentified MNO data needs to evolve into a data pipeline, providing updates on population movement dynamics on a monthly or bi-monthly basis.

LESSONS LEARNED AND RECOMMENDATIONS

This project demonstrated a strong use case for how data for development projects can be deployed. This technical report provides evidence for DIAL's working hypothesis that analytics from MNO data can be used to complement other traditional datasets, such as census data, in order to inform policy and decision-making among governments and their agencies. Insights on population density and movement patterns might also be used in future projects to inform decision-makers about other key locations such as water points, schools, and agricultural cooperatives, in order to improve people's access to those critical services.

The process of acquiring various datasets, analyzing them, and integrating them into country systems generated a number of lessons of findings, limitations, lessons learned, and recommendations. The Project Team hopes these lessons are useful for governments, private sector companies, NGOs, and other implementors, as well as other actors in the digital ecosystem who might be considering similar work.

Key Findings

1. **An estimated 7.74 million Malawians, or 44.7%, live more than 5 km walking distance from a clinic or health post.** There is a trade-off between prioritizing access to the maximum number of people and access to the highest percentage of people per district.
2. **Observations from the deidentified MNO data analytics suggest that there is significant population movement on weekends and during the rainy season.** On weekends, Malawians leave the cities for the lakeshore and market towns. During the rainy season, Malawians migrate from the central region to the south, possibly driven by demand for agricultural labor.[22]
3. **The number of people serviced per new health post, a measure of allocation efficiency, varies significantly across districts.** When allowing for seasonal migration, this efficiency metric shifts, suggesting that the needs for health coverage in a given location may vary across the year. When incorporating the potential for flooding, certain districts are far more vulnerable due to reduced access to health posts. In certain districts, flooding may mean that each health post would have to service an additional 30,000 people.

[22] This is an observation derived from the work the Project Team was doing alongside the government in the region.

Limitations

1. **The model is gender blind by construct**, since the data was stripped of identifying characteristics. 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, so this model likely over-represents the movements of men compared to women, who are more likely to visit health facilities, particularly for pre-natal care (Rowntree, O., 2019).
2. **Certain populations in other vulnerable groups such as the poor, the elderly, and children might also have different rates of mobile use.** Similar to the gender-gap in mobile ownership, people who are members of vulnerable groups likely also have lower rates of mobile ownership. Therefore, the models in this project are likely to under-represent the movements of people who are members of these vulnerable groups.
3. **The model is based on data from one of two principal telecom providers** in the country. The Project Team's validation exercise has found no evidence that this systematically biases the data.
4. **Nighttime location in deidentified MNO data is assumed to indicate where a person lives.** However, this assumption might not properly account for nighttime or shift workers, who might have alternative work schedules outside of 'normal' daytime hours. As a result, this assumption could impact the Project Team's understanding of commuting patterns and short- and medium-term population movements.
5. **There is an unknown composition of cell phone users relative to those who tend to migrate,** and therefore short and medium-term population movements may be either overstated or understated.
6. **By construct, there is no MNO data for the approximately 5% of zones outside of mobile coverage** in Malawi. Population movement and growth in these zones has been inferred based on observed patterns in adjacent areas with mobile coverage, but since these are the most remote areas, these inferences might not provide a complete picture.

Recommendations

1. **Define a specific, demand-driven use case** to structure the analytical process within a realistic time horizon that allows for unforeseen delays.
2. **Bring together a broad-based analytical team of researchers** to tackle the many technical challenges inherent in preparing, cleaning, and analyzing multiple datasets and bringing them together to deliver relevant insights.
3. **Emphasize country-level buy-in from the very beginning of the project** to ensure that the use case is policy-relevant and that the research is in full regulatory compliance with regards to data encryption and user confidentiality.
4. **Engage with private-sector partners throughout the process** on the potential value-add from both a CSR and business development perspective. MNOs are looking to engage with development partners to expand the use of their data products, but expectations must be managed to allow for differing perspectives.
5. **Work continuously with technical counterparts** to ensure the relevance of analytical products, laying the groundwork for integrating these products into country systems.
6. **Consider the country-specific data limitations** and make appropriate corresponding modifications to the assumptions and algorithms to adjust for limitations, to the extent possible.
7. **Validate the model using robust census data,** in order to have high confidence that the MNO data can be used as a good proxy for population density and movement in that particular country. However, in countries that do not have good census (or other) static population data, the model might not be able to work effectively and be properly validated.
8. **Verify the performance** of the model through qualitative evaluation.

Overall, this project has demonstrated that it is possible to bring together public and private datasets to generate usable data analytics which inform public policy decision-making in health resource allocation. It has also proven the usefulness of location-based mobile data in capturing usable insights for government decision-makers, when used together with more traditional census data. Although developing these partnerships is time-consuming, the benefits may be replicable once such relationships are formed. DIAL will be continuing to evaluate the program, including testing the veracity of the modelling and its assumptions, in the coming months and years.

Although developing these partnerships is time-consuming, the benefits may be replicable once such relationships are formed.

Nevertheless, more pilots are needed to understand the relevance of this approach to modeling health resource allocation, as well as to be able to make statements about the validity, return on investment, and ethical considerations of these types of applications of private sector data. To this end, DIAL will continue to work with both government partners in Malawi and other country governments, as well as a range of international and regional stakeholders, to explore how such data for development models can be made replicated across sectors and geographies.

BIBLIOGRAPHY

Anselin, L., and Rey, S. (1991) Properties of tests for spatial dependence in linear regression models. *Geographical Analysis, 23(2)*, 112-131.

Blondel, V.D., Esch, M., Chan, C., Clerot F., Deville, P., Huens, E., Morlot, F.E., Smoreda, Z. , and Ziemlick, C. (2013) Data for Development: the D4D Challenge on Mobile Phone Data. *arXiv.org, Cornell University.* Retrieved from https://arxiv.org/pdf/1210.0137.pdf.

Blumenstock, J., Cadamuro, G., and On, R. (2015) Predicting Poverty and Wealth From Mobile Phone Metadata," *Science, 350(6264)*, 1073-1076.

de Montjoye, Y.A., et al. (2019) On the privacy-conscientious use of mobile phone data. *Scientific Data 5(180286)*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6289108/.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., and Tatem, A. J. (2014) Dynamic Population Mapping Using Mobile Phone Data. *Proceedings of the National Academy of Sciences, 111(45)*, 15888-15893.

Erbach-Schoenberg, E., Alegana, V. A., Sorichetta, A., Linard, C., Lourenço, C., Ruktanonchai, N. W., and Tatem, A. J. (2016) Dynamic Denominators: The Impact of Seasonally Varying Population Numbers on Disease Incidence Estimates, *Population Health Metrics, 14(35)*.

Martinez-Cesena, E.A., Mancarella, P., Ndiaye, M., Schläpfer, M. (2015) Using mobile phone data for electricity infrastructure planning, *arXiv.org, Cornell University*. Retrieved from https://arxiv.org/abs/1504.03899.

Noriega-Campero, A., Rutherford, A , Lederman, O., de Montjoye, Y.A., Pentland, A. (2018) Mapping the privacy-utility tradeoff in mobile phone data for development. *arXiv.org, Cornell Universit*y. Retrieved from https://arxiv.org/pdf/1808.00160.pdf.

Odland, J. (1988) *Spatial Autocorrelation*. Newbury Park, California: Sage Publications.

Pohjankukka, J., Pahikkala, T., Nevalainen, P., and Heikkonen, J. (2017) Estimating the Prediction Performance of Spatial Models via Spatial K-fold Cross Validation. *International Journal of Geographical Information Science*, 1-19.

Ricciato, F., Widhalm, P., Craglia, M., Pantisano, F. (2015) *Estimating population density distribution from network-based mobile phone data*. Report EUR 27361 EN. Retrieved from http://publications.jrc.ec.europa.eu/repository/bitstream/JRC96568/lb-na-27361-en-n.pdf.

Rowntree, O. (2019) *Connected Women the Mobile Gender Gap Report*. Retrieved from GSMA website: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/03/GSMA-Connected-Women-The-Mobile-Gender-Gap-Report-2019.pdf.

Smith, C., Mashhadi, A., and Capra, L. (2013) Ubiquitous sensing for mapping poverty in developing countries. Paper submitted to the Data For Development (D4d) Challenge at Net Mob 2013. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.408.9095&rep=rep1&type=pdf.

Steele, J. E., et al. (2017) Mapping poverty using mobile phone and satellite data." *Journal of The Royal Society Interface 14(127): 20160690*. Retrieved from https://royalsocietypublishing.org/doi/10.1098/rsif.2016.0690.

Stevens, F.R., Gaughan, A.E., Linard C., Tatem A.J. (2015) Disaggregating Census Data for Population Mapping Using Random Forests With Remotely-Sensed and Ancillary Data. *PLoS ONE, 10(2): e0107042*. Retrieved from https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0107042&type=printable.

Tomsu, K., Eggermont, A., and Snel, N.(2016) Telecom data for efficient malaria interventions. *arXiv.org, Cornell University*. Retrieved from https://arxiv.org/abs/1609.08747.

Wesolowski, A., et al. (2014) Commentary: Containing the Ebola Outbreak—the Potential and Challenge of Mobile Network Data. *PLoS Currents*, 1. Retrieved from http://currents.plos.org/outbreaks/index.html%3Fp=42561.html.

Wesolowski, A., O'meara, W. P., Tatem, A. J., Ndege, S., Eagle, N., and Buckee, C.O. (2015) Quantifying the Impact of Accessibility on Preventive Healthcare in Sub-Saharan Africa Using Mobile Phone Data. *Epidemiology, 26(2)*, 223-228. Retrieved from https://insights.ovid.com/crossref?an=00001648-201503000-00015.

WorldPop (2018) *The Spatial Distribution of Population in 2015, Malawi*. Retrieved from https://dx.doi.org/10.5258/SOTON/WP00538.

Xin, L., Wrathalld,D.J., Sundsøye, P. R., Nadiruzzaman, M., Wettera, E., Iqbale, A., Qureshie, T., Tatema, A., Canrighte, G., Engø-Monsen, K., Bengtsson, L. (2016) Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh. *Global Environmental Change*, 38, 1-7.


GOGO CHATINKHA BANDA
MATERNITY WING
BUILT IN 1980
sobo
ONCOLOGY WARD
NJANE
CLINIC

dial Digital Impact Alliance

digitalimpactalliance.org